AD-A057 440 ILLINOIS UNIV AT URBANA-CHAMPAIGN SAVOY AVIATION RES--ETC F/G 5/9
JUDGMENT EVALUATION AND INSTRUCTION IN CIVIL PILOT TRAINING.(U)
DEC 77 R S JENSEN, R A BENEL DOT-FA77WA-3920

UNCLASSIFIED FAA-RD-78-24 NL

1 OF 2
ADA
057440

(12) LEVEL II

Report No. FAA-RD-78-24

JUDGMENT EVALUATION AND INSTRUCTION IN CIVIL PILOT TRAINING

R.S. Jensen R.A. Benel
Aviation Research Laboratory
University of Illinois - Willard Airport
Savoy, Illinois 61820

December 1977
Final Report

Document is available to the U.S. public through the National Technical Information Service, Springfield, Virginia 22161.

Prepared for

U.S. DEPARTMENT OF TRANSPORTATION
FEDERAL AVIATION ADMINISTRATION
Systems Research & Development Service
Washington, D.C. 20590

NOTICE

This document is disseminated under the sponsorship of the Department of Transportation in the interest of information exchange. The United States government assumes no liability for its contents or use thereof.

12 152p.

Technical Report Documentation Page

19

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
18 FAA-RD-78-24		

4. Title and Subtitle	5. Report Date
6 Judgment Evaluation and Instruction in Civil Pilot Training ,	11 Dec[redacted]977
	6. Performing Organization Code
	8. Performing Organization Report No.
7. Author(s)	
10 R. S. Jensen R. A. Benel	
9. Performing Organization Name and Address	10. Work Unit No. (TRAIS)
Aviation Research Laboratory at Urbana-Champaign University of Illinois - ~~Willard Airport~~ Savoy, Il~~linois~~ 61820 406 171	11. Contract or Grant No. 15 DOT-FA77WA-3920 new
12. Sponsoring Agency Name and Address	13. Type of Report and Period Covered
Federal Aviation Administration Development Section C, ALG-313 800 Independence Ave. S. W. Washington, D.C. 20591	9 Final Rep[redacted]rt. [redacted] Nov[redacted]76-Dec[redacted]
	14. Sponsoring Agency Code ~~ALG-313~~

15. Supplementary Notes

16. Abstract Accident statistics reveal that approximately 50 percent of the civil aviation fatalities are in part related to poor flying judgment. What is meant by good flying judgment? Is it "professionalism?" "Maturity?" "Flying experience?" Is judgment something pilots are born with or can it be taught or modified by a flight instructor? How can you tell if a pilot has good judgment?

This report presents an indepth examination of each of these questions from the perspectives of aviation and psychology. A definition of pilot judgment is presented consisting of an intellective part (How well can you think?) and a motivative part (are you cautious or risky?). Evidence was found from research in other fields such as medicine and business that indicates that both aspects of judgment can be taught. This and other research also indicates that judgment can be evaluated.

Combining these research findings with educational principles and the opinions of professional aviation educators, a broad outline for a judgment training and evaluation program is presented. Suggestions are made for implementation of judgment training and evaluation techniques in ground school and aircraft training. For schools with computer-aided instruction programs and/or simulators, some methods for using these devices are suggested. In view of the accident statistics and the favorable findings concerning judgment training in other fields, it is suggested that judgment training could be highly beneficial to civil aviation safety.

17. Key Words	18. Distribution Statement
Judgment Decision Making Pilot Training	Document is available to the U.S. public through the National Technical Information Service, Springfield, Virginia 22161.

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
UNCLASSIFIED	UNCLASSIFIED	152	

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized i

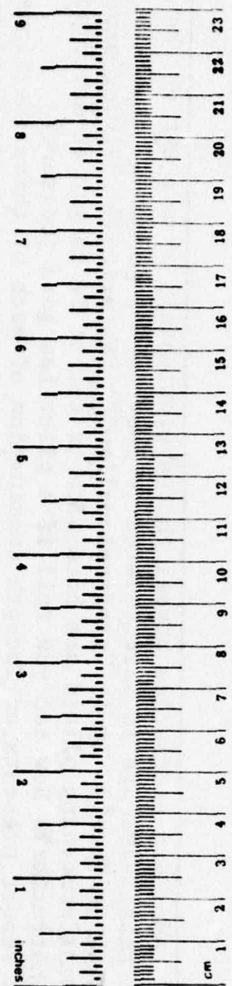
Page i 406 171 Gm

METRIC CONVERSION FACTORS

Approximate Conversions to Metric Measures

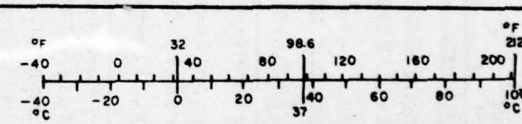
Symbol	When You Know	Multiply by	To Find	Symbol
		LENGTH		
in	inches	*2.5	centimeters	cm
ft	feet	30	centimeters	cm
yd	yards	0.9	meters	m
mi	miles	1.6	kilometers	km
		AREA		
in^2	square inches	6.5	square centimeters	cm^2
ft^2	square feet	0.09	square meters	m^2
yd^2	square yards	0.8	square meters	m^2
mi^2	square miles	2.6	square kilometers	km^2
	acres	0.4	hectares	ha
		MASS (weight)		
oz	ounces	28	grams	g
lb	pounds	0.45	kilograms	kg
	short tons (2000 lb)	0.9	tonnes	t
		VOLUME		
tsp	teaspoons	5	milliliters	ml
Tbsp	tablespoons	15	milliliters	ml
fl oz	fluid ounces	30	milliliters	ml
c	cups	0.24	liters	l
pt	pints	0.47	liters	l
qt	quarts	0.95	liters	l
gal	gallons	3.8	liters	l
ft^3	cubic feet	0.03	cubic meters	m^3
yd^3	cubic yards	0.76	cubic meters	m^3
		TEMPERATURE (exact)		
°F	Fahrenheit temperature	5/9 (after subtracting 32)	Celsius temperature	°C

*1 in = 2.54 (exactly). For other exact conversions and more detailed tables, see NBS Misc. Publ. 286, Units of Weights and Measures, Price $2.25, SD Catalog No. C13.10:286.



Approximate Conversions from Metric Measures

Symbol	When You Know	Multiply by	To Find	Symbol
		LENGTH		
mm	millimeters	0.04	inches	in
cm	centimeters	0.4	inches	in
m	meters	3.3	feet	ft
m	meters	1.1	yards	yd
km	kilometers	0.6	miles	mi
		AREA		
cm^2	square centimeters	0.16	square inches	in^2
m^2	square meters	1.2	square yards	yd^2
km^2	square kilometers	0.4	square miles	mi^2
ha	hectares (10,000 m^2)	2.5	acres	
		MASS (weight)		
g	grams	0.035	ounces	oz
kg	kilograms	2.2	pounds	lb
t	tonnes (1000 kg)	1.1	short tons	
		VOLUME		
ml	milliliters	0.03	fluid ounces	fl oz
l	liters	2.1	pints	pt
l	liters	1.06	quarts	qt
l	liters	0.26	gallons	gal
m^3	cubic meters	35	cubic feet	ft^3
m^3	cubic meters	1.3	cubic yards	yd^3
		TEMPERATURE (exact)		
°C	Celsius temperature	9/5 (then add 32)	Fahrenheit temperature	°F



FURTHER RESEARCH

As indicated in the Executive Summary of this report, the successful completion of a scientific endeavor such as this requires the completion of six steps as follows:

1. Discovery of the need or problem;
2. Statement of the problem in its simplest terms;
3. Observation of relevant factors in the context of the problem;
4. Establishment of hypotheses suggested by these observations;
5. Experimental testing of these hypotheses in the context of relevant factors;
6. Generalization of the experimental findings to real-world problems in the form of rule changes, new hardware designs, or other system modifications.

It was the purpose of the present study to complete steps one through four for the problem of judgment training and evaluation in Civil Aviation. Steps five and six remain for further research and implementation.

In the approach to the problem of the experimental testing of pilot judgment training and evaluation, a necessary step will be the development of specific training techniques for application to the various media available. It is suggested that the initial effort should be concentrated on changes to the ground school curriculum because this medium is available to all, and it is quite amenable to cognitive instruction techniques. Three other training media, the simulator, computer-aided instruction, and the aircraft should also be given consideration in judgment training and evaluation, even though each has limitations in terms of either availability or training value.

It is suggested that changes to the training syllabus take the form of a special section of the training curriculum. Specific judgment training efforts using situational techniques, role playing, balance sheet procedures and other methods mentioned in this report should be used during this section of the course. It is expected that situational techniques will also be used interspersed throughout all flight and ground training as well as using the context of the specific training endeavor.

Because situational techniques appear to be the most effective for judgment training and evaluation, a major initial effort should be to develop a reservoir of judgment demanding scenarios. Some strategies for the development of these scenarios are suggested in the report and others could be developed in further research.

A major part of the implementation of judgment training and evaluation is the training of instructors and evaluators to assume these roles. Because of the nature of judgment training material, the instructor must assume a significant role in the training process. Most instructors are not equipped to communicate this type of material effectively. Therefore, any further research program must examine strategies for training instructors and evaluators in terms of pilot judgment.

It is suggested that psychological tests, some of which were mentioned in the report, should be used as an aid to the development of individual training strategies and as a means of confirming other judgment evaluation findings. These must be a part of any further empirical study of pilot judgment.

Before suggestions for rule changes can be made, there must be a thorough experimental examination of alternative training and evaluation methods. Such a program must include adequate control groups even at the risk of providing less than optimum training for these groups. Measures of the effects of judgment training must be geared to the training received to determine its effect. Measures must also be made of the transfer to the real-world flying environment in terms of the number of accidents and incidents in which participants were involved.

As suggested in the report, a major initial research effort will involve the definition of evaluation criteria that match the training techniques and identify real-world individual needs for further training. The initial step in this process is the development of behavioral objectives. Further study of those suggested in this report and the identification of others is needed.

Although empirical research is needed before rule changes can be made, enough is known at the present time to suggest the following changes:

1. Pilot judgment training and evaluation should be emphasized in instructor and safety seminars using what is currently known.

2. Instructors should be encouraged to develop their own training evaluation strategies and suggest these to the FAA.

3. Psychological tests should be administered to first time candidates for the airman's medical examination in an effort to develop a profile of problem personalities in later flight.

PREFACE

This final report presents the results of an investigation to establish a definition of pilot judgment and to determine whether and how pilot judgment may be taught and evaluated. The contract was supported by the Systems Research and Development Service, Federal Aviation Administration. Patrick Russell is the Contracting Officer's Representative. G. T. Connors of the Logistics Service, Federal Aviation Administration, is the Contracting Officer.

Assistance in business matters has been rendered by John H. Kamerer, Assistant Director of Business Affairs and William Morgan, Contract Negotiator, Research Grants and Contract Division and by John M. Johnson, Business Manager of the Institute of Aviation, University of Illinois at Urbana-Champaign. Assistance in the formulation of ideas for this report was given by many psychologists, flight instructors, and leaders in the aviation training community. The authors wish to thank in particular Robert Durst, Michael Kelley, James Finnegan, and Professor Stanley Roscoe of the Department of Psychology, University of Illinois and Dirk E. Van Dam of Moody Aviation, Elizabethton, Tennessee for their assistance. The authors also wish to thank Cindy Scalamonti for her able assistance in the typing of this manuscript.

EXECUTIVE SUMMARY

The successful completion of any scientific or engineering endeavor depends upon the imagination with which investigators follow a step-by-step process of scientific inquiry (Roscoe, 1971). The steps in this process may be summarized as follows:

1. Discovery of the need or problem;
2. Statement of the problem in its simplest terms;
3. Observation of relevant factors in the context of the problem;
4. Establishment of hypotheses suggested by these observations;
5. Experimental testing of these hypotheses in the context of relevant factors;
6. Generalization of the experimental findings to real-world problems in the form of rule changes, new hardware designs, or other system modifications.

It is extremely important to complete steps one through five of this process before attempting step six, particularly where human safety is involved.

This document is intended as an initial step in the overall investigation and solution of the problem of pilot judgment training and evaluation in Civil Aviation. In essence, the scope of this study is covered by steps one through four given above. Steps five and six await the efforts of further investigation. Our purpose was to reduce the magnitude of this extremely broad problem to a level that is manageable at the empirical level.

The statement of work around which this study is based outlines three broad topics for investigation:

1. A definition of pilot judgment considering its mental, moral, emotional, social, psychological, philosophical, and educational implications;
2. A determination of whether and how judgment can be imparted to the human; and
3. A determination of whether and how judgment can be evaluated both for empirical investigation and for field testing.

Although each of these topics demands a definitive answer, none was anticipated nor was any found because time and financial limitations prevented the conduct of empirical investigations. The methods used in this

initial investigation of pilot judgment included: a literature review, a survey of pilot and flight instructor opinions, and interviews with psychologists and educators. Each of these sources of information proved to be useful to some extent but less helpful than one would have hoped.

For example, the literature review led to many studies related to pilot judgment in a peripheral way but only to one which was directly related (Thorpe, Martin, Edwards, and Eddowes, 1976). A vast literature was discovered in the areas of decision making and automobile safety from the standpoint of decision making, but few offered applications to aviation. However, because decision making in general is closely related to our concept of pilot judgment, these studies are useful to our investigation, and we have pointed out aviation applications for their findings.

The survey of pilots and flight instructors was helpful in that it revealed the importance and complexity of the problem. The formal survey of nine pilot training organizations was less than successful because only two returned written responses. These are given in the Appendix to this report. In formal discussions with pilots and flight instructors at the University of Illinois Institute of Aviation, most agreed that pilot judgment could and must be taught. They further said that pilot judgment is being taught by the actions of the flight instructor to decision situations. Nevertheless, they admitted that it is not being taught in any systematic way. On the subject of evaluation, only one admitted to have failed to recommend a student purely on the basis of poor judgment. However, many said that they could recognize poor judgment in a student. In such cases, because of the lack of a clear definition of judgment, students were failed on the basis of a borderline performance of some other more clearly defined flying maneuver involving skilled performance.

Interviews with psychologists and educators were useful in that they pointed toward relevant literature and psychological tests. They were also helpful in formulating ideas about judgment and judgment modification. In general, psychologists and educators did not feel qualified to speculate about the success one might find in trying to teach judgment to pilots. They did indicate that human judgment, in general, is difficult to modify in adults.

A major part of the investigation concerned the establishment of the extent of the role played by judgment in flight operations and accidents in Civil Aviation. Discussions with flight instructors indicated that an element of judgment is involved in every maneuver performed. Whether the maneuver is taxiing, landing, stall training, or turns-about-a-point, circumstances may exist in any specific performance requiring a high level of judgment (or "planning ahead" or "headwork") to avoid an unsafe operation.

Airline pilots have indicated that a high degree of judgment is involved in their routine operations. Two airline captains representing the Airline Pilot's Association, pointed out in a discussion at the 1977 Annual Meeting of the Human Factors Society that pilot judgment is frequently used in pilot decisions to go beyond previously established rules or standards for reasons of economy, noise reduction, and safety. They stated that, if pilots were forced to follow the rules rigidly, the airlines would be

brought to their knees economically in a very short time. The extent and justification for the use of judgment prerogatives by pilots are not unlike those exercised by air traffic controllers who have frequently demonstrated the effect on the system when rules are followed explicitly.

A third area which dramatically demonstrates the extent to which pilot judgment is used in Civil Aviation is in the assignment of accident causes by the NTSB. Our analysis of the fatal accidents investigated between 1970 and 1974 reveals that 80 percent were in part caused by pilot error. Of these fatal pilot error accidents, over 50 percent were the result of faulty pilot judgment. Although some may disagree with many NTSB analyses, these statistics tend to indicate that the use of pilot judgment in Civil Aviation is extensive and that it is frequently faulty leading to hazardous results.

The results of this investigation are summarized below in terms of the three major requirements of the program: definition, training, and evaluation. As stated above, these results are based on "soft" data (expert opinion and related literature) and should not be considered to be definitive conclusions. However, they are a useful and a necessary prerequisite to empirical study and application to Civil Aviation problems in the area of pilot judgment.

Judgment Definition

Our definition of judgment in flying airplanes has two components:

1. The ability to search for and establish the relevance of all available information regarding a situation, to specify alternative courses of action, and to determine expected outcomes from each alternative.

2. The motivation to choose and authoritatively execute a suitable course of action within the time frame permitted by the situation.

Where:

1. "Suitable" is an alternative consistent with societal norms.

2. Action includes no action, some action, or action to seek more information.

The first part of the definition refers to intellectual abilities. It depends upon human capabilities to sense, store, retrieve, and integrate information. This function is what Van Dam (see Appendix) calls the "discriminating ability" in professional pilots. In signal detection theory, it is called detectability (d'). It is purely rational and could be stated mathematically. If it were possible to separate this part of human judgment from the second part (which it is not), the mind would solve problems in much the same way as a computer. This is not to say that it would be

error free; it deals with probabilistic information. This part is dependent on the amount, type, and accuracy of information stored as well as inherent and learned capabilities to process information.

The second part of the definition refers to motivational tendencies. The emphasis is on the directional aspects of motivation rather than the aspects of motivation dealing with intensity. It says that a part of human judgment is based upon bias factors (costs and payoffs) or tendencies to use less than rational information (defined by society) in choosing courses of action. Society would probably consider the use of any information other than that required to define the safety risk (e.g. monetary gain, gain in self-esteem, adventure seeking, etc.) as less than rational. This part of human judgment is called the response bias (β) in signal detection theory. It is what Van Dam (see Appendix) has called the "response pattern" of the professional pilot. If properly developed, this part of human judgment would tend to halt the use of information not directly related to the safety of the flight and to direct the pilot's decision toward the use of rational processes.

This concept of judgment can be represented as a judgment continuum based on cognitive complexity and decision time. At one end of the continuum are what might be called cognitive judgments. As described above, these judgments are very complex in that they usually involve a large number of relevant pieces of highly probablistic information, they usually require the specification of and choice from among several alternatives, and they are frequently affected by emotions, values, and social pressures. In addition, cognitive judgments usually permit some deliberation before a control response is required.

At the other end of the judgment continuum (not considered in this report) are the more common perceptual judgments of distance, altitude, speed and clearance. Perceptual judgments are less complex in that they involve fewer pieces (frequently one) of fairly accurate information from which responses are made with highly learned motor behavior. They may require simple responses but frequently call for immediate control movement.

Judgment Training

A number of people have questioned whether or not pilot judgment could be taught at all. To answer this question and to take a preliminary step toward a judgment training methodology we used the opinions of flight instructors and other experts in aviation training and a review of pertinent literature. This investigation has led us to believe that although basic personality characteristics are unlikely to be changed very much by any kind of pilot training, flying judgment, in terms of learned discriminations and response patterns, can and must be taught in aviation instruction programs. Flight instructors tend to agree with the assertion made in the Aviation Instructor's Handbook (1977) that basic flying attitudes and safety practices, such as following checklists, adhering to loading limitations, and observance of weather minimums, can and are being taught by example if not systematically by flight instructors today.

The safety record of the airline industry provides evidence that one type of judgment training, i.e. memorizing and following set procedures for every possible emergency, are partially effective. A new research program in the Air Force is seeking to improve these "Boldface" procedures with "Situational Emergency Training" (Thorpe, et al., 1976). They suggest that discrimination abilities and response selection tendencies can be taught with a minimum of memorization of procedures, and that such training would lead to better pilot judgment. The basic technique is to expose the student pilot to many and varied flight situations which require a discrimination of relevant stimuli from among many situational dimensions and to make responses in these situations. Van Dam (Appendix) suggests the use of similar techniques for teaching judgment in "borderline" situations.

The research literature on human decision making further suggests that both discrimination abilities and response selection tendencies can be taught or modified. Goldberg (1968) discovered that "value" training techniques were effective in teaching discriminative abilities for clinical psychological judgments. Models of mental processes used by expert medical diagnosticians and livestock judges have been used in training people to discriminate effectively in these fields (Shanteau and Phelps, 1977; Hoffman, Slovic, and Rorer, 1968).

Basic research has indicated that response selection tendencies also can be modified. The research on signal detection theory has provided ample evidence that human response tendencies or biases can be readily manipulated from very cautious to very risky depending on the cost and payoffs associated with the various outcomes (Lee, 1971). Janis and Mann (1977) using a "conflict-theory" model of decision making have developed and tested several new procedures to improve decision making under the titles of "awareness-of-rationalizations, emotional role-playing, and balance sheet." They report that these procedures have demonstrated effectiveness in changing decision-making tendencies and in attitude modification.

A systematic approach to judgment training is suggested emphasizing the development of training objectives, precisely controlled learning experiences to achieve these objectives, development of training and transfer criteria, evaluation of training and transfer effectiveness, feedback of evaluation information to improve the training program, and recognition of interactions among components within the system. A model of an instructional system is presented consisting of three phases: need-assessment, curriculum development and training, and evaluation. All three phases are needed to accomplish the goals of a systematic approach to any instructional program. The feedback from the evaluation phase to the need-assessment phase indicates that an instructional system is never complete. It is particularly important to any new program such as that proposed for judgment training to provide and use evaluation information as the basis for adjustments throughout the system.

Situational training techniques are suggested as the basic elements of judgment training. Such techniques are applicable to four training media: ground school (lectures, movies, discussions, role-playing, etc.), computer-aided instruction, simulator instruction, and aircraft instruction.

Although all four of these training media are not typically available in pilot training programs, training that is conducted in media that are available should include a significant segment devoted to pilot judgment.

Because situational training places high demands upon instructor creativity and communication capabilities, the initial phase of any such program must be devoted to the development of these skills in the instructors themselves. As suggested by the Air Force SET program, one part of the instructor training course could be devoted to the development of a pool of hypothetical emergency or borderline scenarios that would permit the systematic manipulation of specific situational dimensions for later use in judgment instruction programs. Of particular importance in these instructor training courses would be training instructors to recognize when students are properly or improperly analyzing a situation in terms of all relevant dimensions. Student evaluation capabilities such as these are not typical among flight instructors today but seem warranted by the importance of proper judgment training and evaluation in Civil Aviation.

Judgment Evaluation

Perhaps the most difficult part of any study of human judgment is the evaluation of performance. The reason is that much of what must be evaluated cannot be observed directly but must be inferred from observation of other related behaviors. In our search for information on judgment evaluation, we sought first to determine how judgment is currently being evaluated (a requirement of all flight examinations) and then to seek other approaches that might be effective. From our discussions with flight instructors and pilot examiners, it is clear that judgment is not being evaluated effectively today. The reason, they suggest, is that, until now there has been no clear definition of pilot judgment for use in such evaluations.

One approach to evaluating pilot judgment is offered by Van Dam (see Appendix). In his approach, the evaluation begins with pretesting using psychological examinations prior to admitting students for flight instruction. Initial impressions from these pretest examinations are combined with other indicators of judgment performance such as "obvious effort and attention to instruction", "relaxation", "division of attention", "response delays", "confidence", "capacity for problem-solving", and "initiative". In later pilot training, evidence of judgment development is seen through an "eagerness to learn or high motivation", "teachability", "adaptability and flexibility", "an intuitive quality in thinking or decision-making", "a pattern of good choices", and "applies margin and allowances". Careful records of student judgment performance against these criteria would be useful as a judgment training device as well.

Evidence from the psychological testing literature indicates that complex judgmental behavior can be evaluated with varying degrees of success in other contexts. The key seems to be the establishment of a structural setting and carefully defined criteria or indices of desired performance. For example, McGuire and Babbott (1967) report that evaluation of medical physician judgments can be made using simulated clinical problems in a carefully structured environment. Scoring was done by a group of experts classifying the judgments into categories from poor to excellent. Levine

and McGuire (1968) demonstrated that judgmental evaluations may be made on the basis of role playing by candidates before a group of medical examiners. They concluded that this technique is more effective and lower in cost than direct observations. In a related study, Levine and McGuire (1970) found that the structuring of oral exams, the standardizing of case materials, and the training of examiners minimizes the problems of evaluation.

Pretraining evaluations of judgment ability in pilot training candidates is a potentially useful adjunct to the entire instructional and evaluation process. Results from such tests could be used by training management to adapt their training program to emphasize training in areas identified as potentially weak in pretraining tests. Tests which identify risk-taking tendencies (Kogan and Wallach, 1964; Taylor and Dunnette, 1974) and tests which identify accident proneness (Shaw and Sichel, 1971) are potentially useful in this regard. A situation specific testing technique developed by Shealy (1973) is potentially useful for pretraining evaluations and for test development for use later in training as well.

The development of clearly defined judgment evaluation criteria presents the greatest challenge to proper evaluation of pilot judgment for all phases of pilot training. To insure the evaluations are made along the same dimensions as the training conducted, the development of these criteria should be based on pre-established behavioral objectives. In this study, positive statements of acceptable pilot judgmental behavior were developed through an examination and integration of NTSB reports of accidents in which pilot decisional behavior was found to be at fault. Similar judgmental criteria could be developed for every major maneuver taught. These could be graded by the instructor together with evaluations of knowledge and skill each time the maneuvers are attempted.

In pilot training, for each level of pilot experience, certain judgmental hurdles (proficiency levels) could be objectively specified. The instructor or examiner who evaluates the judgments would have a range of acceptable performances, also objectively specified. Evaluation of pilot judgment would be a matter of comparing performance against the established criteria in a carefully structured situation. The critical point for judgment evaluation in a national system is the use of the same criteria by all judges as well as by the pilots themselves.

The definition of pilot judgment has two components: discrimination among situational dimensions and response selection. Both components must be evaluated. To operationalize these components for use in any specific training or testing situation, the evaluator may ask the following questions:

1. <u>For discriminative judgment</u>: Did the student consider all of the alternatives available to him? Did he consider all of the relevant information and assign proper weights to each? Did he integrate the relevant information efficiently before making his choice?

2. <u>For response selection tendencies</u>: Did the student exhibit any tendency to consider factors other than safety (such as his own self-esteem, adventure, or social pressure)

in making his response selection? Did he seem to be highly prone to use semi-relevant factors such as financial gain and convenience in situations where safety should have been the primary consideration?

As suggested above, the use of such criteria as these requires more of the evaluator than just an occasional passing glance at the instrument panel. It requires the careful structuring of the situation, perhaps hypothetically, and a careful examination of actions taken by the student. It probably would require a dialogue between the student and the evaluator to establish what the student actually considered in making his choice. Each evaluation must be considered a training device as well, and as such, feedback should be given to the student concerning all aspects of the decision situation. We recognize that evaluations of this sort place high demands on the flight instructor. Nevertheless, they seem warranted in view of the high number of fatalities caused by faulty judgment which is hardly being evaluated at all under the present system.

CONTENTS

LIST OF FIGURES

GLOSSARY

adjunct - something joined or added to another thing; a valuable quality or attribute.

amenable - liable to be brought to account or judgment; capable of submission (as to a judgment or a test).

a posteriori - of or relating to the kind of reasoning that derives propositions from the observation of facts or that by generalization of facts derives principles; proved by induction from facts obtained by observation or experiment.

a priori - by reasoning from definitions formed or principles assumed; without examination or analysis; independently of experience.

arbitrarily - arrived at in an arbitrary manner (arbitrary - arising from unrestrained exercise of free will, caprice, or personal preference).

asymptotic - from asymptote - (in geometry) a straight line associated with a curve such that as a point moves along the curve, the distance between the curve and line approaches zero and the slope of the curve approaches the slope of the line; (in example p.22 this report, no longer changing significantly).

attenuating - (from attenuate) to lessen the amount, value, force of; to make less complex, to reduce the severity of.

attrition - breaking down or wearing down from repeated attacks or constant diminution; gradual loss of strength (in the example p.95, the loss of numbers through the selection procedures and training dropouts).

cognitive - of, relating to, being, or involving the process of knowing.

concomitantly - in a manner occurring along with, at the same time as, and with or without causal relationship.

contingencies - events or conditions occurring by chance and without intent, viewed as possible and eventually probable, depending on uncertain occurrences and coincidences.

cursory - rapidly often superficially performed with attention to detail.

deleterious - hurtful, destructive, noxious, pernicious.

dichotomous - dividing into two parts or groups; (in the example p.98, capable of being reduced to two part decision, e.g.: yes-no, go-no go, etc.).

empirical - originating in or relying or based on factual information, observation or direct sense experience; capable of being confirmed, verified or disproved by observation or experiment.

epistemological - relating to or based on a theory of knowledge (roughly how we go about knowing what we know).

espoused - adopted as a matter of policy; believed in, furthered, supported, or defended.

euphoria - feeling of well-being or elation; especially one that is groundless or disproportionate to its cause or inappropriate to one's life situation.

hedonistic - characterized by a way of life that asserts that pleasure or happiness is the sole or chief good in life.

hypothetical - of or depending on supposition; conditional, assumed, conjectural.

indices - (plural of index) somethings that serve as pointers or indicators.

induced - to cause, to act as by powers of persuasion, to bring on or bring about.

inference - the act of passing from one or more propositions, statements, or judgments considered as true to another truth which is believed to follow from the former.

innate - existing from birth; belonging to the essential nature of something; inborn, hereditary, inherited.

intangibility - (from intangible) incapable of being touched or perceived by touch; incapable of being defined with a certainty or precision.

integrative - favoring or implementing the process of forming into whole by addition or combination of parts or elements.

intellective - relating to or based on the power or faculty of knowing.

judicious - having or exercising sound judgment; characterized by discretion.

mentalistic - of or relating to mental phenomenon (introspectionistic - relating to examination of one's own thoughts and feelings).

milieu - environment; setting.

motivational - of or related to process of motivating (stimulating the active interest in); drive; incentive.

neurosis - psychological disorder characterized by anxiety, phobias, obsessions or compulsions.

operationalize - to define concepts and terms used in nonanalytic scientific statements in terms of identifiable and repeatable operations (observable events).

postulate - to assume; claim as true, existent, or necessary.

probabilistic - based on a theory that certainly is impossible and probability suffices to govern belief and action.

prohibitive - restraining from desired course or action.

psychodrama - a procedure in which a hypothetical real-life situation is acted out.

psychometrics - technique of psychology applying to use of mental measurement.

psychophysical - interrelating the physical and psychic; pertaining primarily to the method developed for studying the perception of physical magnitudes.

psychosis - profound psychological disorder manifested by disorders of perceition or thinking.

purport - to convey, imply or profess outwardly.

ramifications - resulting development; consequence.

scenarios - plot outlines (especially hypothetical situations provided with open-ended consequences).

situational dimensions - relevant aspects of a situation which might be varied (e.g., the conditions of the aircraft, weather, etc.).

subsumed - viewed, listed, or rated as a component in an overall or more comprehensive classification.

viable - affecting the imagination, senses, or mind as real, genuine, artistically whole, or important; capable of being put into practice.

JUDGMENT EVALUATION AND INSTRUCTION IN CIVIL PILOT TRAINING

BACKGROUND

The increasing complexity of flight operations and airborne systems combined with the increasing demands in our society for safety, dependability, economy, and reduced energy consumption, in recent years, have placed mounting pressures on the aviation community. Technological advances adopted by one sector of the aviation community have affected all other sectors; their benefits and demands are not felt exclusively by the aeronautically sophisticated. These technological advances have created demands for new levels of skill, knowledge, and judgment to which few pilots have been trained, and the training costs to prepare civilian pilots to operate safely and effectively in the changing system are becoming prohibitive. Furthermore, there is inadequate assurance that all those operating aircraft in the system are currently qualified to do so; this problem is growing.

Flying has developed so rapidly that there has been little time for a serious study of what flying is all about, particularly in terms of how pilots think. Many changes in regulations are expedients designed to solve a problem that has already developed. Often solutions to existing problems create new problems, which in turn are "solved" by new regulations. The

problem of training new pilots and retraining current pilots to facilitate the implementation of new procedures and regulations in a mobile but energy limited society is just beginning to be recognized.

However, if it were merely the teaching of flying skills and the training of pilots to follow changing procedures and regulations, the problem of training and retraining pilots to operate safely in our complex aviation system would be a much smaller one than it is. Unfortunately, because actual conditions are never quite the same as those used to develop aviation regulations and procedures, the safety of a given flight also depends upon a significant amount of interpretation and evaluation of existing conditions by the pilot.

Rational Pilot Judgment

For example, the conditions used to develop flight performance values for a particular type of airplane may be ideal including clean airplane surfaces, a new engine, a new propeller, an unrestricted air filter, and a company test pilot. In actual conditions, the pilot must compare the performance values provided in the book for these ideal conditions with those in which he finds himself. These actual conditions may include a dirty airplane, a slightly used engine, a few marks on the propeller, a slightly dirty air filter, and a less than perfect pilot. He must then look at many other conditions such as gross weight, center of gravity, wind, temperature, humidity, etc. for comparison with those used in the book to determine his expected flight performance. Finally, he must check the present and forecast weather, the terrain, expected traffic density,

and compare them with an estimate of his own capability before making a decision on whether his planned flight will be safe.

Examples such as these requiring pilot judgment with less than perfect information are available in all areas of flight activity. In addition, every decision that the pilot makes is colored by physiological, psychological, and social pressures that are virtually impossible to weigh properly on the spot. For example, just as persons watching a sporting event may "see" an infraction or foul differently depending upon their vantage points and which team they support, a pilot may be influenced to view the weather outlook and his own abilities differently depending on the importance or value he assigns to a given flight. The person's self-image and his need to maintain his external image largely determine how much effect different values or rewards for making a flight will have on his judgment of his ability to make a safe flight (Kogan and Wallach, 1964). The willingness of the pilot to never exceed legal limitations, the ability of the pilot to evaluate all conditions affecting the safety of a given flight, and the willingness of the pilot to extend safety margins accordingly are measures of pilot judgment that must be moderated by the pilot's education.

Irrational Pilot Judgment

Whereas the above discussion deals with the rational aspects of pilot judgment, because pilots are human, they are susceptible to social pressures and inherent mental weaknesses that may result in less than rational pilot judgment representing the other end of a continuum. Such irrational pilot judgment is responsible for such unsafe flying practices as flying under

bridges, landing on busy highways, attempting to land in football stadiums, and flying "formation" on other aircraft flown by unsuspecting pilots. There are many potential sources of social pressure which, combined with certain inherent mental weaknesses, may lead to these types of activities.

These pressures include peer reactions, fear of failure, censure from superiors or family members, and many others. Individual susceptibility to these pressures may be difficult to modify through education, but they may be identified through some testing procedures. It is a difficult task to test individuals not only for their knowledge, skill, and rational judgment capabilities, but also for their irrational judgment tendencies as these apply to safe flying. However, it is apparent from accident statistics (given below) that new approaches to pilot testing are necessary to improve the safety of civil aviation.

Functional Model of Man-Machine System

Before discussing some possible improvements to pilot training and testing, it is necessary to examine functionally the pilot/aircraft man-machine system and to establish the role of pilot training in the development of these functions. A functional model of this system is presented in Figure 1. The cerebral shaped central portion of this diagram presents the functions performed by man, and the functions in the remainder of the diagram are performed by the machine (aircraft). The functions of man can be associated, somewhat arbitrarily, with three activity categories: procedural, decisional, and perceptual-motor. Procedural and perceptual-motor activities have both cognitive and motor aspects and include sensing,

transforming, recollecting, recognizing, and manipulating functions. Decision making is purely cognitive and consists of choosing courses of action from the perceived alternatives. Action choices are mediated by procedural and perceptual-motor activities (Roscoe, 1974). The role of pilot training then is to provide the knowledge necessary for the recollecting function, to develop the skill necessary for all procedural and perceptual-motor activities, and to develop the deciding function so that the pilot will make consistent and safe aviation decisions.

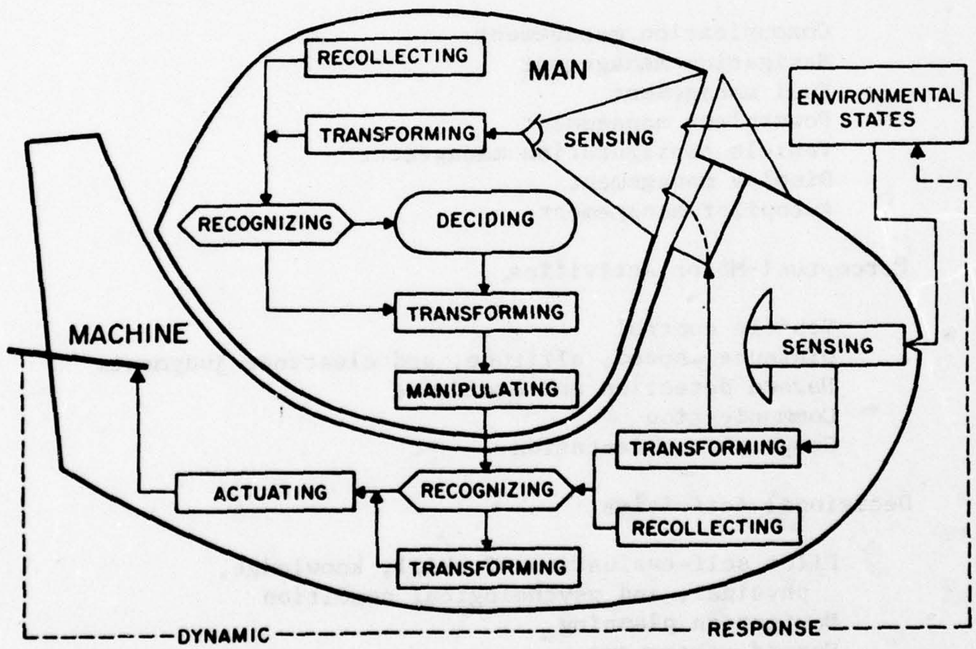


Figure 1. A functional model of man-machine operations (after Roscoe, 1974).

Training Objectives

A determination of an effective pilot training system including airplane, ground simulator, and/or auxiliary ground instruction should start with a categorical analysis of training objectives associated with the end product - a pilot licensed to fly under a certain set of regulations. Civilian training objectives may be classified under a set of behavioral categories useful in specifying associated ground training system characteristics:

Procedural Activities

- Communication management
- Navigation management
- Fuel management
- Powerplant management
- Vehicle configuration management
- Display management
- Autopilot management

Perceptual-Motor Activities

- Vehicle control
- Distance, speed, altitude, and clearance judgments
- Hazard detection and avoidance
- Communication
- Geographic orientation

Decisional Activities

- Pilot self-evaluation of skill, knowledge, physical, and psychological condition
- Navigation planning
- Hazard assessment
- Assessment of attention requirements
- Assessment of aircraft and ground system capabilities
- Mission priority adjustment

Each of the training objectives for specific flight operations can be classified under one or more of these behavioral categories. Consideration

of the types of training for which simulators have demonstrated capability yields the sudden realization that, although they have proven most effective for teaching perceptual-motor skills (Flexman, Roscoe, Williams, and Williges, 1972), simulators in primary pilot training even in the military have been virtually unused in the teaching of decisional skills. The airlines, which have used simulators successfully in teaching decision making, have restricted training to experienced pilots.

The fact that simulators have been used little in teaching decision-making skills in primary training is not surprising in view of the intangibility of such skills and the difficulty of defining good decision-making performance and judging when it has occurred. Nevertheless, few would argue that decisional activities distinguish the effective aircraft commander from the ineffective. In spite of the fact that the development of these skills was heretofore largely left to flight experience, the cultivation of decisional skills or judgment is an instructional objective calling for situational training that may best be carried out with safety only in a simulated flight environment, and to a perhaps lesser but still effective extent by auxiliary training aids.

Effectiveness of Current Pilot Training Programs

One way to examine the effectiveness of current pilot training (and maintenance) programs is to examine the safety record of various modes of transportation. Preliminary 1976 data indicate that the airline safety record has improved to 0.0002 fatalities per million passenger miles. Although the general aviation safety record is not as good, it too has

improved, with 0.2765 fatalities per million vehicle miles in 1976. For comparison, the automobile safety record of 0.0331 fatalities per million vehicle miles in 1976 was intermediate between commercial and general aviation. It is safer to drive then to fly yourself or with friends, but it is far safer than either to fly on our domestic airlines.

Another way to examine the effectiveness of current pilot training programs that may help to identify weaknesses is to analyze general aviation accident data in which pilots were "found to be a contributing cause or factor." Statistics from the National Transportation Safety Board (NTSB) Automated Aircraft Accident and Incident Information System from 1970 through 1974 were used in this analysis. Pilot cause/factors from the NTSB data were classified into the three behavioral categories (procedural, perceptual-motor, or decisional activities) given above. Then the total numbers of both fatal and nonfatal accidents during the five-year period were determined for each of these behavioral categories. The results of these analyses are shown in Table 1.

Although these statistics of pilot caused accidents reflect the influence of more factors than just pilot training deficiencies alone, examinations of these date provide valuable indications of possible weaknesses in current training programs. A classification such as that shown in Table 1 may be somewhat arbitrary because, in many cases, cause/factors may have been listed under more than one behavioral category, and others do not fit very well under any of the categories. Therefore, this classification should be considered preliminary.

Table 1. Number of fatal and nonfatal general aviation accidents in which the pilot in command is listed as the cause or a factor for all data between 1970 and 1974 for the three behavioral categories.

Procedural Activities

		Five Year Totals	
		Fatal	Nonfatal
1.	Failed to extend landing gear	1	255
2.	Failed to retract landing gear	4	14
3.	Failed to use or incorrectly used miscellaneous equipment	14	62
4.	Improper IFR operation	110	66
5.	Improper fuel management	105	1231
6.	Improper starting procedure	1	30
7.	Failed to assure gear down and locked	1	207
8.	Misused or failed to use flaps	27	235
9.	Inadvertently retracted landing gear	0	104
10.	Retracted gear prematurely	1	26
	Total for Procedural Activities	264	2230
	Percent of total pilot caused accidents	4.6	8.6

Perceptual-Motor Activities

		Fatal	Nonfatal
1.	Delayed action in aborting takeoff	5	236
2.	Delayed in initiating goaround	32	380
3.	Failed to see and avoid other aircraft	128	196
4.	Failed to see and avoid object	166	757
5.	Failed to maintain flying speed	846	1825
6.	Misjudged distance, speed, altitude, clearance	351	2864
7.	Failed to maintain adequate rotor RPM	16	153
8.	Improper operation of powerplant controls	53	685
9.	Improper operation of brakes/flight controls	1	688

Table 1, continued

		Fatal	Nonfatal
10.	Improper operation of flight controls	164	569
11.	Improper leveloff	10	1596
12.	Improper compensation for wind	12	550
13.	Control interference	0	1
14.	Improper recovery from bounced landing	5	811
15.	Spatial disorientation	528	60
16.	Failure to maintain directional control	13	1978
17.	Premature liftoff	11	302
18.	Failed to abort takeoff	26	257
19.	Failed to initiate goaround	8	637
20.	Exceeded design stress limits of aircraft	121	16
	Total for Perceptual-Motor Activities	2496	14561
	Percent of total pilot caused accidents	43.8	56.3

Decisional Activities

		Fatal	Nonfatal
1.	Operation of aircraft with known deficiencies	84	201
2.	Operation beyond experience/ability	170	368
3.	Continued VFR into known adverse weather	717	343
4.	Continued flight into known severe turbulence	18	7
5.	Improper inflight decisions/planning	236	597
6.	Exercised poor judgment	235	767
7.	Operated carelessly	7	38
8.	Selected unsuitable terrain	22	1230
9.	Initiated flight into adverse weather	124	61
10.	Psychological condition	11	4
11.	Selected wrong runway	11	341
12.	Failed to follow approved procedures	145	425
13.	Inadequate preflight planning or preparation	511	2341
14.	Lack of familiarity with aircraft	121	611
15.	Started without proper assistance	6	89
16.	Became lost/disoriented	68	248

Table 1, continued

		Fatal	Nonfatal
17.	Taxied, parked without proper assistance	0	67
18.	Left aircraft unattended	1	8
19.	Diverted attention from operation of aircraft	111	501
20.	Inadequate supervision of flight	62	610
21.	Spontaneous improper action	15	119
22.	Misunderstood orders/instructions	3	20
23.	Incapacitation	50	7
24.	Physical impairment	203	65
25.	Inadequate training	0	5
26.	Direct entry	9	14
	Total for Decisional Activities	2940	9087
	Percent of total pilot caused accidents	51.6	35.1

Nevertheless, some useful information can be gained from an examination of the statistics as presented. For example, a majority of the non-fatal pilot-caused accidents (56.3 percent) were the result of faulty perceptual-motor behavior. The most significant factors here, failure to maintain flying speed and misjudgment of distance, speed, altitude, or clearance, are included in one aspect of pilot judgment. On the other hand, a majority of the fatal pilot-caused accidents (51.6 percent) were the result of faulty decisional behavior. The most significant factors in this case were the familiar "continued VFR into known adverse weather" and "inadequate preflight planning or preparation." It is quite apparent from these statistics that these are areas that require additional attention in the training and testing of pilots.

In evaluating the effect of faulty pilot judgment on general aviation accident statistics one must consider two aspects of the deciding function, perhaps representing two ends of a continuum. The first is the general decision process which requires of the pilot a thorough evaluation of the sensed data based on his recollections of previous experiences (see Figure 1). Included in this aspect of the deciding function are all items listed under "decisional activities." The second involves responses at the perceptual-motor level. In this case information is sensed, transformed, and recognized. Then, because of time constraints, a thorough evaluation is bypassed, and a "decision" to manipulate the controls is made. Included here are distance, speed, altitude, and clearance judgments. It is apparent from the accident statistics that both aspects of the deciding function are important to safe flight and possibly suffer from neglect in the present training process.

Subsequent evidence will show that each student brings to the pilot training situation a flexible but preset decision tendency or judgment capability. This tendency, which may be characterized by attitudes including self-esteem, priorities, general intelligence, and perhaps even phobias, contributes to the decision process in the operational flying situation. The in-flight decision process is further complicated by the fact that flying is an emotional experience for many people. High levels of emotion, whether in normal or emergency flight situations, can be expected to have extreme effects on rational decisions, either adaptive or maladaptive. The pilot who has the ability to rank flight alternatives objectively in order of merit and the will to act accordingly in all flight situations may be said to possess good flying judgment.

The Judgment Definition Problem

In spite of all that has been said about it and in spite of many years of experience using the term in aviation training, operations, and accident analyses, very little is commonly known about pilot judgment. Undoubtedly, there are many reasons for this lack of common knowledge, but perhaps the most important is that the study of pilot judgment largely involves an analysis of mental processes, an area that was intentionally avoided by psychologists between 1900 and 1960 because of the tremendous influence of behaviorism as espoused by J. B. Watson and B. F. Skinner. Since 1960, cognitive psychology, which has been willing to examine mental states and processes, has witnessed a rebirth and has grown to rival behaviorism as an influential force in psychological research.

Although a large amount of research has been done on various aspects of judgment in recent years, no one has specifically examined the judgment problem faced by the pilot, the flight instructor, and the pilot examiner. There appear to be three major problems that require solution before major improvements to pilot training and evaluation can be realized in this area. The first is the establishment of a common definition of judgment as it applies to flying. At present, even though the term is used repeatedly in aviation circles, no such definition exists, and the term means different things to different people.

The second major problem is to determine whether pilot judgment can be taught at all. Some consider judgment innate; some believe that it is learned in one's early years and cannot be modified thereafter; others are convinced that judgment, like any behavior, can be modified at any age. Obviously, the final answer to this problem must await empirical investigation. However, there is a vast array of data showing that some aspects of judgmental behavior can be modified. Damos (1977) has recently shown that the time-sharing of attention, an ability generally thought to contribute to good judgments in complex situations, not only can be taught but is transferable from one dual-task situation to others having nothing in common with the first except the requirement for divided attention.

The third major problem is to determine whether pilot judgment can be evaluated reliably, meaningfully, and objectively. If judgment is purely a mental process, it may be difficult to evaluate in any reliable way. On the other hand, behavioral events frequently have been used as indicators of mental activity. Although personality tests have proved

to be somewhat unreliable, these instruments may be useful in evaluating and predicting judgmental behavior. Nevertheless, final answers to this problem also require empirical investigation and analysis of current training, testing, and operational procedures. General recommendations will be made for changes to current procedures based on the preliminary solutions to the problems mentioned above and an analysis of these procedures.

One is reminded of the old negro spiritual: "Ever-abody talk about heaven ain't gonna heaven." So many people in aviation talk about judgment, but they rarely define what they mean. Definitions, however, are arbitrary conveniences - neither true nor false. It is the privilege of any theorist to establish his own definitions, hopeful that his readers will find them not discordant with their own thinking and of equal convenience (Berrien, 1976). There is probably no greater obstacle to proper judgmental training and evaluation in aviation than the lack of a common operationally oriented definition of the concept.

It has been pointed out that judgment has been used to describe a variety of things in aviation. Perhaps its most common usage has been to describe the mental activity that takes place as a pilot develops a perception of his clearance between or over objects, his altitude near the ground, his speed over the ground, or his distance to an object such as the runway. A second, less common, usage of the term judgment describes the mental activity involved in choosing a course of action from among alternatives. Obviously, this second usage of the term is similar to the first in that both involve making choices.

However, there is a basic difference. The first refers to highly learned perceptual responses that must be made in a very short time, in some cases continually. The second refers to cognitive decisions for which set procedures have not been established or may have been forgotten. Usually, more time is available to evaluate the situation; usually a larger number of possible courses of action must be considered, and there is a greater degree of uncertainty concerning the existing situation and possible outcomes than in the case of perceptual judgments. For these reasons, cognitive judgments have been the source of greater misunderstanding in pilot training and evaluation.

There are many aspects of cognitive judgment involving nearly all areas of psychology, some areas of philosophy, and perhaps even some areas of theology (Lee, 1971). Before offering a definition of "judgment" as it applies to aviation, a number of factors to be considered in arriving at a definition will be discussed. These may be classified as: mental, moral, emotional, social, psychological, and philosophical factors.

The Mental Factor. The mental factor refers to knowledge, intelligence, information, processing skills, and recall abilities as they affect one's capability for rational judgment. This factor contributes to the quality of the rational or mathematical aspects of the decision process found in the informational search and evaluation. Consideration of the mental factor in judgment opens up the whole area of subjective probability and mathematical decision theory (Lee, 1971; Edwards and Tversky, 1967). Although it is not

intended to be directly applicable to cognitive judgments, the following discussion of the theory of signal detection (TSD) illustrates the two aspects of the judgment definition to be presented later.

Within the past quarter century, TSD (Birdsall, 1955) has been developed into one of the most dominant influences in psychophysics. More importantly, because of the generality of its basic assumptions, certain aspects of TSD are finding increased application in such diverse areas as perception, learning, and human memory. Since TSD was not specifically designed to address itself to the decision-making or judgmental aspects of the pilot's task, it seems appropriate to dissect the pilot's task into pieces small enough for TSD to digest in its own fashion. Following this dissection the appropriate step would appear to be a resynthesis of the task to its original state.

According to traditional TSD, each change in the physical energy at a sense organ elicits a corresponding change within the observer. However, the new stimulus arrives against a background of ongoing stimulation. The probability for the detection of a stimulus event is a function of the intensity of the stimulus relative to the background. Both the stimulus (signal) and background (noise) vary in magnitude, and the frequencies of the varying magnitudes are generally represented by two Gaussian distributions (see Figure 2). One of the most important contributions of TSD is the provision for dividing the decision behavior of an individual into two components representing the observer's sensitivity and his response criterion or bias.

The sensitivity component depends mainly on the physical parameters of the stimuli in the situation and the sensory apparatus of the observer. For simplicity, an observer's sensitivity is represented by d' in Figure 2. Technically, d' is the difference between the means of the distributions divided by the common standard deviation of the distributions. The more distinct the two distributions become, and hence the greater value of d', the more likely is the observer to detect the presence of the signal. In Figure 2, the two distributions overlap: therefore, it is impossible for the observer to respond with certainty whether or not a signal has occurred. We can assume the signal and noise generate central neural effects that, for purposes of decision making, may be summarized by a single quantity. The observer adopts various decision strategies that depend a great deal on other than stimulus parameters.

Despite being independent of sensitivity, the second component of decision behavior also influences discriminative behavior. Because the value of d' is normally considered a constant, variations in signal detections are attributed to such variables as motivation, knowledge of the signal's probability of occurrence, and the costs and payoffs attendant with a given response. Collectively these affect the bias of the observer labelled β. The detection task is essentially a comparison of the generated central neural activity located along the x-axis (decision axis) of Figure 2 with β also located along this axis.

TSD approaches this problem as a statistical decision-making task (Wald, 1950). Specifically, the observer responds "signal" whenever the value of the generated central neural event equals or exceeds the value

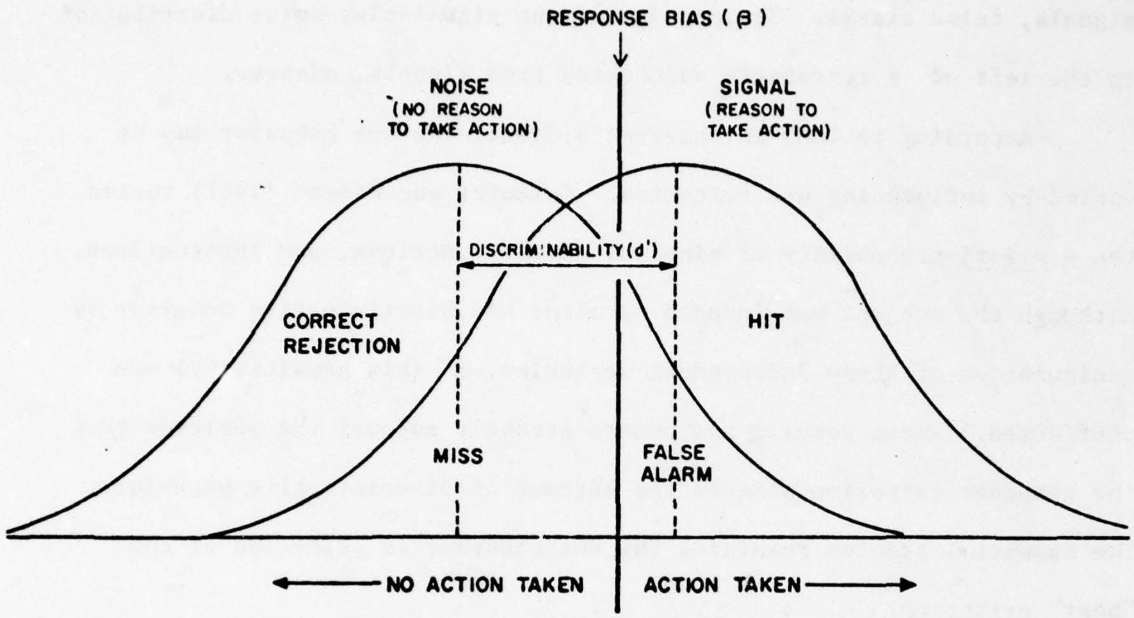


Figure 2. Two Gaussian distributions representing noise and signal in TSD discriminability model.

of β and "no signal" whenever it is less. With overlapping distributions, the observer can vary β and influence the success and failure of his decision-making behavior. Figure 2 represents two possible states of the world, noise alone or signal plus noise, and the two possible responses, action or no action. The observer has two chances for being correct (correct rejection and hits) and two chances for error (misses and false alarms). Using the observer's personal values for hits and false alarms, the value of β can be calculated. In Figure 2, with β set at the indicated position, the observer will correctly identify many true signals and also correctly reject many instances of noise without signals -- hits

and correct rejections respectively. Unfortunately, the portion of the noise distribution to the right of β represents noise identified as signals, false alarms. The portion of the signal-plus-noise distribution to the left of β represents undetected true signals, misses.

According to TSD, an observer's discriminative behavior may be varied by influencing his criterion. Galanter and Holman (1967) varied the *a priori* probability of signals, payoff functions, and instructions. Although the subject was induced to alter his discriminative behavior by manipulation of these independent variables, d' (his sensitivity) was unaffected. These results and others strongly support the position that the response criterion affects the outcome of discriminative behavior. The essential problem remaining for the observer is selection of the "best" criterion.

Birdsall (1955) advanced a definition, the Expected Value Observer, with particular relevance. Each of the four possible outcomes from the experiment (see Figure 3) has a value to the subject; V_H - value of a hit, V_{CR} - value of a correct rejection, V_{FA} - value of a false alarm, V_M - value of a miss. The chosen criterion is that which maximizes the expected value:

$$\beta = \frac{P(N)}{P(SN)} \cdot \frac{V_{CR} - V_{FA}}{V_H - V_M}$$

Optimum performance does not mean errorless performance, and optimality will vary among individuals just as the various values can cause shifts within an individual. β selection maximizes the return for that individual at the time of selection. Shelly and Bryan (1964, p.9) propose that optimum be considered loosely as the "... selection of an action that is in some sense 'best'." With the personal, subjective nature of the concept, it can

	NOISE (NO REASON TO TAKE ACTION)	SIGNAL (REASON TO TAKE ACTION)
NOISE (NO ACTION TAKEN)	CORRECT REJECTION	FALSE ALARM
SIGNAL (ACTION TAKEN)	MISS	HIT

Figure 3. Payoff (cost) matrix representing possible states of the world in TSD paradigm.

be seen that the particular β used by a subject will be entirely consistent with his value structure at the time of selection.

In the aviator's environment the situations in which TSD might potentially be applied are manifold. In each aspect of the pilot's task, TSD or a modification might be appropriately applied from the planning phase, through physically determining the airworthiness of the aircraft, during preflight inspection, to enroute decision making. Some of the more classical applications involve the detection of signals in the preflighting of the aircraft to determine its airworthiness. The inspection of each item on the preflight walk-around provides information that must be compared with an internal representation of the item. Discovery of a discrepancy (for example, a crack on a control surface) is an event typical of TSD tasks.

The criterion that the pilot sets for this situation may vary depending on the relative importance of the flight. It would be possible to take vernier caliper readings of a crack and call all those over a certain length "signals", but even that would not eliminate a human element from the measurement process. When reading engine instruments we could expect to see similar results, for example, in judging the acceptability of the drop in manifold pressure when checking magnetos. Would the criterion of the pilot vary with the costs and payoffs? If a businessman pilot is making a flight to seal a deal that will avert imminent bankruptcy for his company, we might expect his criterion to be somewhat more liberal than if he were making a routine call on a small account. Implicit within all such situations is the concept of a potentially, objectively verifiable signal. Also implicit is the likelihood that "signal" responses will vary, dependent on the personal situation of the respondent.

Although d' is generally taken to be a fixed value reflecting signal strength and the individual's sensory apparatus, some improvement in this area is admissible. It is implicitly assumed that the receiver in the TSD model is well aware of the characteristics of the signal. To the extent that these characteristics may be learned, we could expect to see improvement in accuracy of detection. Savage (1954) in his coverage of decision theory touches on this approach to certainty through experience. He indicates that the person typically becomes almost certain of truth as his experience increases indefinitely.

Presumably for the certificated pilot an asymptotic level of certainty has been approached. During the initial training of private pilots, basic inexperience bears on the decision process. Chernoff and Moses (1959) highlight the role of experience in decision processes. In a common sense

approach, the more experience one has with the various consequences of different courses of action, the more likely the correct alternative will be selected. The above examples indicate how the combination of β and d' can account for decision making in the relatively objective examples given.

A non-trivial fraction of the fatal aircraft accidents each year result from incorrect perceptions or recollections of objectively verifiable data, such as misjudging speed, misunderstanding or forgetting clearance, etc. However, the majority of "important" decisions arrived at in the aircraft environment are based on probabilistic information. The application of classical TSD measures to decisions arrived at in the dynamic, probabilistic, and personal world of the pilot has not been previously demonstrated. Remembering the flexible nature of β, it is easy to see why decisions to continue a flight "against the odds" frequently occur.

These could be attributed to the pilot's over estimating his own ability, to external pressures causing selection of a relatively conservative β for warning signals. Although the nature of these external circumstances would vary among situations, they would be amenable to experimental determination. Types of possible pressures might include social, emotional, and also financial, as in the earlier example of the businessman pilot. Although this approach takes some liberties with the basic ideas of TSD, it is not completely foreign. Its major departure is the change in task from the simple detection to a global decision task.

Although criticism of this approach could be expected in view of the traditional application of TSD, a more liberal application serves to allow a wider range of data to be subsumed. Additionally, a less traditional conception of the pilot's decision task allows investigation of behavior

that might previously have been considered too complex or perhaps mentalistic. The unobservable cognitive behavior of deciding does have observable behavior tied to it, and the influence of decision from the observed behavior should not unduly stretch the imagination.

The Moral Factor. The moral factor refers to the influence of ethical values on pilot judgment. Ethical values are dispositions to behave in certain ways because one believes that he ought to behave in those ways. These dispositions are developed early in life but are continually modified by one's environment. Philosophers say that the moral value of a judgment is the moral difference between a choice made by an individual and the choice that society would have made in the same situation. The moral factor tends to influence the choice of a course of action in the direction believed to be the societal norm. Based on our definition of good pilot judgment, this influence should be considered positively.

There is strong evidence indicating that the immediate context can have a considerable effect on moral judgment. In a study reported by Parducci (1968), a "test" of moral judgment was given to a lecture class of several hundred college undergraduates. The procedures were described as follows:

> Each student was asked to rate the moral value of different acts of behavior in terms of his "own personal set of values." His task was to assign each act to one of five categories: "1—not particularly bad or wrong," "2--undesirable, a good person would not do this," "3--wrong, highly questionable," "4--seriously wrong," and "5--extremely evil." Half of the students were given a list made up mainly of relatively mild acts of wrongdoing; the other half got a nastier list, consisting principally of acts that could be counted on to evoke strong disapproval.

> The crucial feature of the experiment was that, embedded among the other items, each list contained six items that were common to both. The students were cautioned to "judge each act just as though it were the only one you were judging. In other words, don't try to compare the acts or rank them while forming your judgments."

In spite of the instructions to judge each item alone, students could not escape the immediate context. All six acts were rated more leniently by those who judged them in the context of the nastier list than they were rated by those who saw them in the milder context. Pocketing the tip left by a previous customer gets a much lower rating in the context of the nasty list (2.46) than it does in the context of the mild list (3.32). Although these data suggest a universal balance for judgments based on context, Parducci provides other data indicating that the distribution of items making up the context has an even greater effect. Therefore, the definition of judgment must consider the context and the distribution of items making up the context.

In the aviation situation, these results suggest several implications. First, the motivation to choose certain courses of action is influenced by the context (societal norms) in which the pilot operates. Second, because the flight instructor is an influential part of the pilot's context, he has the opportunity and responsibility to influence his student's response bias in the direction of societal norms. Judgment can be taught to some limited extent. Third, because passengers (and flight students) expect pilots to make decisions based on societal norms (or even more cautiously), it is the responsibility of the pilot to know these norms and to bias his decisions in their direction.

The Emotional Factor. The emotional factor refers to the effects of emotional arousal in the form of fear, worry, stress, anxiety, euphoria, etc. on pilot judgment. Moderate levels of emotional arousal are required for optimum decisional performance, but extremely high levels have either a beneficial or a detrimental effect on judgmental processes. Worry and euphoria are especially bad for decisional activities because, even at relatively low levels, they divert attention from the decisional task to the individual himself (Wine, 1971; Liebert and Morris, 1967). Stress may result from concern over material or social losses one might suffer from whichever course of action is chosen -- including the costs of failing to live up to prior commitments (Janis and Mann, 1976).

Stress may also result from expected loss of self-esteem as a competent decision-maker. Stress resulting from fear of physical danger has some effect on decision-making capability but probably not as much effect as the other above mentioned causes of stress. The emotional factor apparently affects decision making in two ways: it may either divert attention from the task at hand resulting in a degrading of discriminability of important stimuli, or it may mobilize attention and thereby lower discriminability thresholds. In either case, it tends to bias the decision in the direction of following through with prior plans and commitments.

The Social Factor. The social factor refers to the effects of social pressure to follow certain courses of action. Laboratory studies have shown that an individual can be induced to conform to the wishes of others

by applying pressures that weaken his confidence in his own beliefs while strengthening his confidence in the correctness of the others' views. In a laboratory study Asch (1951) found that subjects conformed to incorrect perceptual judgments in a group setting. When other members of the group were unanimous, subjects conformed even in the face of obvious evidence to the contrary.

This finding supports what is normally referred to as social influence. Deutsch and Gerard (1955) divided social influence into normative and informational types which they felt were relevant to Asch's findings. The distinction is that normative applies to an influence to conform to another's expectations, while informational applies to influences to accept information from others as evidence about reality. In an unreported experiment, Simon (personal communication) found that ground control intercept (GCI) radar observers "detected" targets at undetectable ranges when they heard another observer (the experimenter's "skill") in the adjacent booth consistently calling targets at unusually long ranges.

It is not surprising that judgments of others (particularly when they are perceived to be motivated and competent to judge accurately) should be taken as evidence to be weighed in coming to one's own judgment. We have learned in the past that others are frequently reliable sources of information and expect them to be so in the future. The effects from these situations would be informational social influence. Once a group (even if loosely structured) is established, normative social

influence assumes greater importance. Thus, the particular structure of the situation can have subtle effects on the outcome of an individual's judgments.

Sources of social pressure that could affect pilot judgment include fellow pilots, family members, air traffic controllers, passengers, superiors, and friends. Effects of social pressure on a pilot who is unsure of his judgment can be great with the stress greatest when the judgment does not conform to that of the group. A very real possibility exists for errors in judgment to occur. Errors biased toward the limit of regulation and man-machine capabilities would appear to be as likely as errors of over-caution and, from the NTSB statistics, of much greater consequence.

The Psychological Factor. The psychological factor refers to all four of the above factors as well as to the effects of personality. There are many aspects of a personality that could conceivably affect judgment. Two that have undergone extensive investigation (Kogan and Wallach, 1964), are test anxiety (motive to avoid or fear of failure) and defensiveness (excessive concern with image maintenance). Kogan and Wallach found that individuals with high levels of test anxiety and defensiveness, both of which are considered to be motivational predispositions, demonstrate a maximum of conservatism in risk-taking tasks. Individuals with lower levels of these two aspects of personality demonstrate greater tendencies to assign a higher priority to task or situational difference. Therefore, these individuals could be expected to evaluate the relevant information more carefully, and they would be more objective in the assignment of likely

outcomes of various alternatives.

Another line of research that may be fruitful in terms of the development of a judgment evaluation instrument for aviation is reported by Shealy (1974). He tested three instruments for their ability to discriminate risk-takers from cautious individuals in skilled task performance. Two of the instruments, the Need for Achievement Test (Atkinson, 1957) and the Locus of Control (the Internal-External Scale devised by Rotter, 1966), are generalized tests. The third, the Situation-Specific Test, was an experimental instrument derived by Shealy for the experimental task, downhill skiing.

The dependent measure in the real-world experiment was bodily injury during a season of skiing. Shealy's results indicated that for real-world situations, situation specific tests are more discriminating than are general, nonspecific tests. In Shealy's judgment, "a successful instrument for discriminating potential accident victims from the population at large was demonstrated." The potential for such a test, specific to flying, is apparent.

<u>The Philosophical Factor.</u> The philosophical factor provides insight into the meaning of judgment by placing three basic restrictions on its definition and interpretation. These restrictions fall into the following categories:

1. Naive realism vs. indirect apprehension of the life world.
2. Relative vs. absolute value systems.
3. The problem of inference.

The first of these may seem to be trivial, but the implications of the theory of knowledge in which one operates are rather staggering, especially if reliable prediction of phenomena is required. A naive realist would claim that the world is exactly as we see it, i.e., there is no question that the world is exactly as our senses tell us it is. This is the framework in which most of us operate on a day to day basis, and its utility is obvious. However, it is well known (especially to psychologists) that humans do make errors in accessing phenomena in their life world. They see things that aren't there and fail to report things that are.

There are well known distortions of perception especially in the visual and auditory fields. Because of such phenomena, most of us begin to think of the human as seeing the world in a special way that is partially determined by the physical structure and operation of sensory inputs and processing, and partially determined by the manner in which we process information, our learned "programs." In essence, the first restraint simply reminds us that we do not experience the world directly; we perceive it through the mechanics of our senses, and then interpret these inputs. Such processes are obviously not immune to error.

The second restriction is critical to the definition of judgment because it establishes the criteria for determining whether a judgment is right or wrong. In a nutshell the problem is choosing between absolute criteria that do not vary with circumstance (i.e., an act is judged wrong regardless of the circumstances) and relative criteria (i.e., ones in some respect dependent on the conditions in which the judgment occurs). There is by no means a hard and fast distinction between the two because absolute

judgments may be deviations from established procedures for specific situations, or, in the *a posteriori* sense, they may be identified strictly with error.

However, if error is the sole criterion of bad judgment, the issue is obscured, for we have merely identified the two phenomena and are now faced with the problem of separating them in our assessment of situations in general. Similarly, if judgments are entirely relative, there is no sense in asking which are good and bad since there is literally no general sense of judgment. The correct path may be a happy medium with absolute judgments applied for standard procedures in specific kinds of situations but always modified by the condition that such fixed procedures are based on the relative judgments of an experienced group or by society in general.

The third restriction is the strictest: the problem of inductive inference. The problem goes back at least to David Hume in the 18th Century and is as annoying today. To put it simply, it is impossible to predict future phenomena from the occurrence of past phenomena with certainty. Just because, within my experience (and that of anyone I've encountered), objects fall when they are dropped, does not allow me to say with complete confidence that a pencil (or any other object) will fall when it is dropped. This is not to say that it won't, and in fact we can say that it probably will, but such an assumption is unstable when based on only two cornerstones: observation of past occurrences of the phenomenon in question and inference that things in the future will behave as they have in the past.

That is, we *assume* that the laws of nature are consistent over time. There is no logical basis for this assumption. It is simply made *ad hoc*,

without even the possibility of proof. It is a jump in the dark. As such, it is a central problem addressed by Persig (1974) in his psychological, philosophical, and instructional novel, Zen and the Art of Motorcycle Maintenance. Persig convincingly "proves" that it is impossible to prove anything, even the generality for all time of the laws of gravity or thermodynamics. Hypotheses can never be proved, only disproved, and at any moment there is an infinity of acceptable hypotheses to explain any phenomenon, each waiting to be disproved.

Now that these three restrictions have been established, let us take a look at how they might affect a definition of judgment. The first restriction tells us that we might be mistaken about the way the world "really" is because our senses are fallible. This of course means that judgment assumes first and foremost an accurate representation of the world. This assumption is tacitly incorporated in any appeal to authority in a given perceptual situation. If we tell someone to react to a situation in a certain fashion, we assume that that person will perceive the situation in a manner at least similar to the way we do. In other words, we assume that humans are much the same in the way they view the world (at least perceptually), but we also recognize that sometimes some of us are out of whack (we don't agree on what is perceived).

The second tells us that, though procedures (absolute values) may be established for specific situations, such procedures are in fact derived from relative value systems in that they are based on the consensus of experts. The application of such procedures to the population involved carries the implied assumption that all members of that population are the

same, physically, perceptually, and ethically. This is clearly not the case, and this in turn means that there may be times when good judgment for an individual involves not following the established procedure. Thus, while perceptual differences (illusions, disorientation) may be coped with by following set procedures, unusual ambient circumstances may dictate improvisation when established procedures become inappropriate. In the latter case, success may be a criterion of good judgment but should not be the only criterion. That such departures from set procedures are made routinely on the basis of the operator's judgment can be seen from the actions of ATC controllers who perform "slow-downs" by following set procedures. Pilots have also indicated that they could bring their companies to economic disaster by following set procedures exactly (Stone at Human Factors Society annual meeting, October 1977).

The last restriction warns us that appropriate procedures in the past may not be appropriate in the future. We are faced with a problem at two levels. First, we must form some theory or set of laws that allows us to predict the future behavior of individual pilots with a high degree of confidence. Such laws are always drawn from past observation, so (1) the accuracy with which such laws describe past phenomena and their interactions, and (2) the assumption that such relations will not vary greatly in the future. The first of these is an empirical problem of "goodness of fit"; the second is beyond our control and involves a "leap of faith."

DEFINITION OF JUDGMENT

The preceding background discussion emphasizes the open endedness of defining judgment. Even if we could determine what correct judgment is, in retrospect there is no way of guaranteeing that the definition would remain the same over time. We are left with at least two possible alternatives: (1) defining good judgment as, following the societally accepted procedure in question when the action taken proves effective and bad judgment as those cases when it does not; and (2) leaving judgment open ended, or purely subjective, which precludes accurate assessment and therefore is of no practical use to anyone.

In practice there may be a happy medium between such extremes. Certain procedures gleaned from consensus of experts could be defined as optimum. Individual performance would be assessed on the basis of appropriate application of these procedures in given situations. At the same time, such procedures should not be considered absolute, thereby stressing the importance of development and use of individual judgment rather than the blind following of procedure absolutely. Thus, a practical definition of judgment in flying airplanes involves two factors:

1. The <u>ability</u> to search for and establish the relevance of all available information regarding a situation, to specify alternative courses of action, and to determine expected outcomes from each alternative.

2. The <u>motivation</u> to choose and authoratively execute a suitable course of action within the time frame permitted by the situation.

Where:

1. "Suitable" is an alternative consistent with societal norms.

2. "Action" includes <u>no</u> action, <u>some</u> action, or action to <u>seek</u> more information.

The first part of the definition is intellective. It deals with mental storage, retrieval, and integrative capabilities. This part is purely rational and can be stated mathematically. If it were possible to separate this part of human judgment from the second part (which it is not), the mind would solve problems in much the same way as a computer. This is not to say that it would be error free; it deals with probablistic information. In addition, this part is dependent on the amount, type, and accuracy of information stored as well as inherent and learned capabilities to process information. This part of the judgment definition is covered by the whole area of subjective probability and mathematical decision theory (Lee, 1971; Edwards and Tversky, 1967; Shanteau, 1975; Slovic and Lichtenstien, 1968).

The second part of the definition is motivative. The emphasis is on the directional aspects of motivation rather than the aspects of motivation dealing with intensity. It says that a part of human judgment is based upon bias factors and/or tendencies to use less than rational information (defined by society) in choosing courses of action. Certain personality characteristics such as fear of failure and defensiveness have been found to bias decision-making in risky situations (Kogan and Wallach, 1964). Social pressure in the immediate context has also been found to bias decision-making (Asch, 1951). Finally, emotional arousal in the form

of fear, worry, stress, anxiety, and euphoria has been found to effect both the mental processing capabilities and level of motivation in human judgment (Wine, 1971; Liebert and Morris, 1967; Steiner, 1972; and Janis and Mann, 1976).

Applying this definition to the art of flying, good pilot judgment requires a thorough prior knowledge of factors affecting the safety of a flight, a capability for separating out information relevant to the safety of a given flight, a capability for specifying alternative courses of action relating to a given flight, and a capability for accurate determination of expected outcomes from each alternative. In addition, good pilot judgment requires a recognition that biasing factors, such as stress or euphoria, can lead to judgmental errors. Thus, the motivational aspect recognizes that biasing factors exist. Good pilot judgment requires the supression of these error producing factors so that decisions can be made on the basis of relevant factors from the physical world.

The major goals of pilot judgmental training should be (1) to improve probabilistic decision-making capabilities and (2) to develop pilot motivation to avoid the adverse effects of decision-biasing and other decision-degenerating factors in flying situations. Conversely, the major thrust of pilot judgmental evaluation should be to measure the level of pilot attainment of these two training goals. The purpose of this report is to present some principles that may have useful applications in both the training and evaluation of pilot judgment.

It should be pointed out, for clarity, that the concept of judgment as we have defined it represents a judgment continuum based on cognitive

complexity and time. At one end of the continuum are what might be called cognitive judgments. As described above, these judgments are very complex in that they usually involve a large number of relevant pieces of highly probabilistic information, they usually require the specification of and choice from among several alternatives, and they are often affected by emotions, values, and social pressures. In addition, cognitive judgments usually permit some deliberation before a control response is required.

At the other end of the judgment continuum are the more common perceptual judgments of distance, altitude, speed, and clearance. Perceptual judgments are less complex in that they involve very few pieces (frequently one) of fairly accurate information that can be responded to with highly learned motor behavior or compliance with established procedures. They may require simple responses but frequently call for immediate control movement as in a conventional tracking task. Although a philsophical argument could be made to the contrary, these perceptual judgments require different teaching and evaluation methods from cognitive judgments, and they are not within the scope of this discussion.

CAN PILOT JUDGMENT BE TAUGHT?

The first question to be addressed following the establishment of the definition is whether or not pilot judgment, as defined, can be modified through training. An examination of pilot training and training research literature gives reason to doubt that judgment can be taught. Literature and syllabi commonly used in flight instructor courses contain large sections on how to teach the motor skills of flying but very little on how to teach pilot judgment (see the FAA's Aviation Instructor's Handbook, 1977). The typical private pilot course contains a scattering of judgmental instruction in the areas of weather avoidance and power-plant emergencies but no systematic judgmental training.

The pilot training research literature has an even worse record in the area of pilot judgment. The research has dealt primarily with the transfer effectiveness of flight simulators in the areas of procedural and perceptual-motor activities, finding that these devices are highly effective in teaching these skills (Flexman, Roscoe, Williams, and Willeges, 1972; Hashbarger, 1967; Povenmire and Roscoe, 1973; Williams and Flexman, 1949; and Caro 1973). However, very little research or use of simulators is evident in the area of pilot judgmental instruction at any level of pilot training. Thus, one might be tempted to conclude that pilot judgment cannot be taught.

However, there is evidence in aviation showing that at least one form of judgmental training, assigning procedures for every conceivable situation that might arise, may be effective. This evidence is in the form of airline data on use of simulators in flight crew training. Demonstrations by American Airlines (Gibson, 1969) and by Trans World Airlines (Trans

World Airlines, 1969) offer convincing support for the conclusion that complex simulators are effective both for the training and testing of pilots. Operational statistics compiled by American Airlines on simulator and airplane time required to reach criterion (performance level adequate for captain transition), shown in Figure 4, reveal the effectiveness of airline simulator instruction for transition training.

Before the introduction of the simulators in 1966, pilots required around 20 hours of airplane time for the left-seat transition in the Boeing 727 and 707. With the introduction of the simulator and its more effective use in recent years, the aircraft time required to reach this criterion was reduced to approximately 1.5 hours by 1974. With the introduction of flight checks on revenue flights, this number has not been effectively reduced to zero. Furthermore, the total hands-on instruction time (simulator and aircraft), by 1974, had been reduced to under 20 hours in both the 727 and the 707.

This very effective use of the simulator in airline applications is due largely to the fact that applicants are screened for decision-making ability both before and during training and that their pilot training programs employ a high degree of decision-making instruction in these devices. Although this judgmental training primarily involves the teaching of procedures for use in every conceivable emergency (as opposed to development of reactive thought processes and reduction of biasing factors), an airline pilot may perceive the consequences of poor decisions (such as, making a flight knowing that one's physiological and/or mental functions are impaired by lack of sleep, medication, use of alcohol, etc.) to be more severe than a private pilot under similar circumstances would believe.

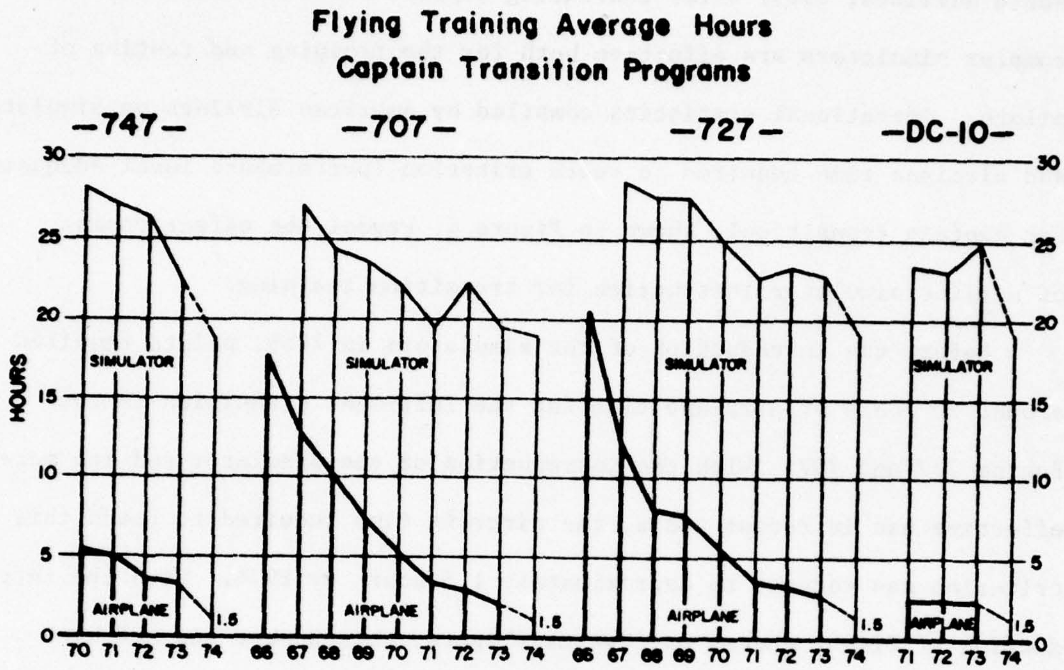


Figure 4. Average simulator and airplane time required in transition to Captain at American Airlines for four types of airplanes.

Whereas, an airline pilot may risk becoming unemployable for a similar job within the industry, a private pilot in addition to a perceived lower probability of being "caught" may expect to risk no more than a suspension or revocation of his certificate. On the other hand, perhaps some of the principles used by airlines in teaching such things as failure analysis procedures might be applied effectively in beginning pilot training programs for both military and general aviation.

Looking outside the field of aviation one finds other evidence indicating that judgment may be taught. For example, although the theory of signal detection (TSD) was not designed specifically to handle cognitive judgments, many of its basic methods can be used to explain and perhaps

even modify pilot-judgment behavior. As explained previously, TSD divides an individual's decision behavior into two components representing his sensitivity (d') and his response criterion or bias (β).

The sensitivity is affected both by the physical value of the stimuli in the situation (signal vs background noise) and the quality of the sensory apparatus of the observer. In cognitive judgment this is the intellective component. On the other hand, the response criterion or bias (β) represents the point in the signal-to-noise distribution at which the observer is willing to say "signal". It is the amount of information needed to tip the decision one way or the other. It is influenced by motivation, knowledge of the signal's probability of occurrence, and the costs and payoffs attendant with a given response. In cognitive judgment the response criterion is the motivative component.

In psychophysical judgments, TSD says that the sensitivity component is quite stable for a given individual. However, the response criterion or bias can be easily manipulated through a wide range of values by adjusting probabilities, costs, and payoffs. (Birdsall, 1955; Wald, 1950; Galanter and Holman, 1967). Although experimental evidence may not prove it, we can infer from the vast amount of psychophysical decision data that cognitive judgments can and are being modified in a similar way. Decision biases, attitude, risk tendencies, consideration for passenger safety, and pilot motivation can and are being taught by the flight instructor by example, if not systematically, at all levels of pilot training. These tendencies are taught, perhaps unconsciously, by the assessment

or assignment of probabilities, costs, and payoffs to actions of the student by the instructor.

There is a growing field of research and literature indicating that sensitivity, or the intellective aspect of cognitive judgment, can be taught as well. For example, attempts have been made to discover the mental processes that are used by expert judges such as stock brokers, livestock judges, and medical diagnosticians in making their decisions (Shanteau and Phelps, 1977; Slovic, 1969; Anderson, 1969; Hoffman, Slovic, and Rorer, 1968; and Hoffman, 1960). The argument is that if models of the mental processes used by expert judges in decision-making were available, these could be used in training people to be expert judges. In each of the areas studied, judgmental training occurs over a fairly long apprenticeship program in which the trainee observes the expert make decisions and learns by observation. However, as in aviation, because of the complexity of the information used to make decisions, observation or even trial and error are very inefficient training methods.

Goldberg (1968) reports an experiment in which various judgmental training methods are used including "outcome feedback," "basic formulas" (for integration of information), and "values" (or coefficients in the formulas, similar to models of mental processes used by expert judges). The task was to learn to make a differential diagnosis of psychosis versus neurosis based on information from a complex personality test. The results revealed that naive subjects (those having no previous psychological training) using the "value" training method could learn to make judgments as accurately as experienced clinical psychologists and in a

much shorter time. Another study reported by Todd and Hammond (1965) found similar results using the Brunswik Lens Model (Brunswik, 1956) as a judgmental training method. These studies indicate that intellective judgment can be taught.

The research on the motivative aspect of cognitive judgment also indicates that training can have a beneficial effect. The major research efforts in this area by Janis and Mann (1977) were mentioned briefly in an earlier discussion. These authors, speaking from a clinical standpoint, begin with the assumption that psychological stress is a frequent cause of errors in decision-making. They say that stress arises from at least two sources:

> First, the decision-maker is concerned about the material and social losses he might suffer from whichever course of action he chooses--including the costs of failing to live up to prior commitments. Second, he recognizes that his reputation and self-esteem as a competent decision-maker are at stake. The more severe the anticipated losses, the greater the stress.

Janis and Mann have constructed a "conflict-theory" model of decision-making postulating that the way we resolve a difficult choice is determined by the presence or absence of three conditions: "awareness of risks involved, hope of finding a better solution, and time available in which to make the decision." They have developed several new procedures to improve decision-making under the titles of "awareness-of-rationalizations, emotional role-playing, balance sheet, and outcome psychodrama. They report that these procedures have demonstrated effectiveness in changing decision-making tendencies and in attitude modification.

Application of such procedures to the flight training situation will be discussed later.

Additional evidence indicating that the motivative aspect of judgment can be modified through training is offered by Pelz (1976). He discusses a technique for reducing overconfidence, recklessness, and poor judgment among young male drivers. The technique involves nondirective discussion of short films of incidents that may confront a young driver as well as group discussion of the young driver's personal highway experiences. This approach reportedly results in a long-term improvement in judgment and accident rates especially among groups of drivers who normally exhibit reckless and overconfident driving behavior.

In summary, a great deal of evidence is available indicating that both intellective and motivative aspects of human judgment can be modified through training. This evidence includes the remarkable record of airline simulator training in which judgmental training primarily consists of the establishment of procedures to handle all expected emergencies. The theory of signal detection demonstrates that judgmental bias can be manipulated through the adjustment of the individual's perception of probabilities, costs, and payoffs. Other studies report the successful use of mental processing information from expert judges to teach individuals to make good intellective judgments. Finally, some research indicates that the motivative aspect of judgment may also be modified through the use of new techniques.

A SYSTEMATIC APPROACH TO PILOT JUDGMENTAL TRAINING

The need for pilot judgmental training has already been established for all levels of flight instruction. The evidence presented in the preceding discussion indicates that pilot judgment can be taught. The questions that remain are: what approaches should be taken to implement pilot judgmental training and what approaches should be used to evaluate the level of judgment possessed by a pilot or flight student. The purpose of this section is to present some answers to the training question. The following section will discuss the evaluation question.

Definitions and Constraints

Before proceeding, it is necessary to establish some definitions and constraints involved in any approach to judgment training. First, training and education, which have been distinguished elsewhere (Glaser, 1962), will be considered equivalent and defined as "the systematic acquisition of skills, rules, concepts, or attitudes that results in improved performance in another environment" (Goldstein, 1974). The "system approach," a term which has been used and abused in many ways, should, in this training context, emphasize the specification of instructional objectives, precisely controlled learning experiences to achieve these objectives, criteria for performance, feedback within the system, and a recognition of the interaction among system components.

In addition to these, the approach to pilot judgmental training should consider the following constraints: the cost and time required of both student and instructor, the qualifications required of the flight instructor and examiner, and the safety requirements to administer such a program. Finally, although a systems approach is used to develop the training context, the major burden of pilot judgmental training falls directly on the flight instructor. He or she is responsible for the creation and use of innovative situational teaching techniques. The following are some guidelines for use in the development of these techniques.

An Instructional Model

A model of an instructional system adapted from one developed by Goldstein is useful for the establishment of the system context for pilot judgmental training. This model shown in Figure 5 presents three basic interrelated phases in a closed-loop instructional system: need-assessment, development and training, and evaluation. All three phases are needed to accomplish the goals of a systematic approach to any instructional program. The feedback from the evaluation phase to the need-assessment phase indicates that an instructional system is never complete. It needs continual adjustment based on the results of the evaluation phase and inputs from the environment.

The Need-Assessment Phase. This phase consists of the establishment of the instructional need and a derivation of behavioral objectives. An

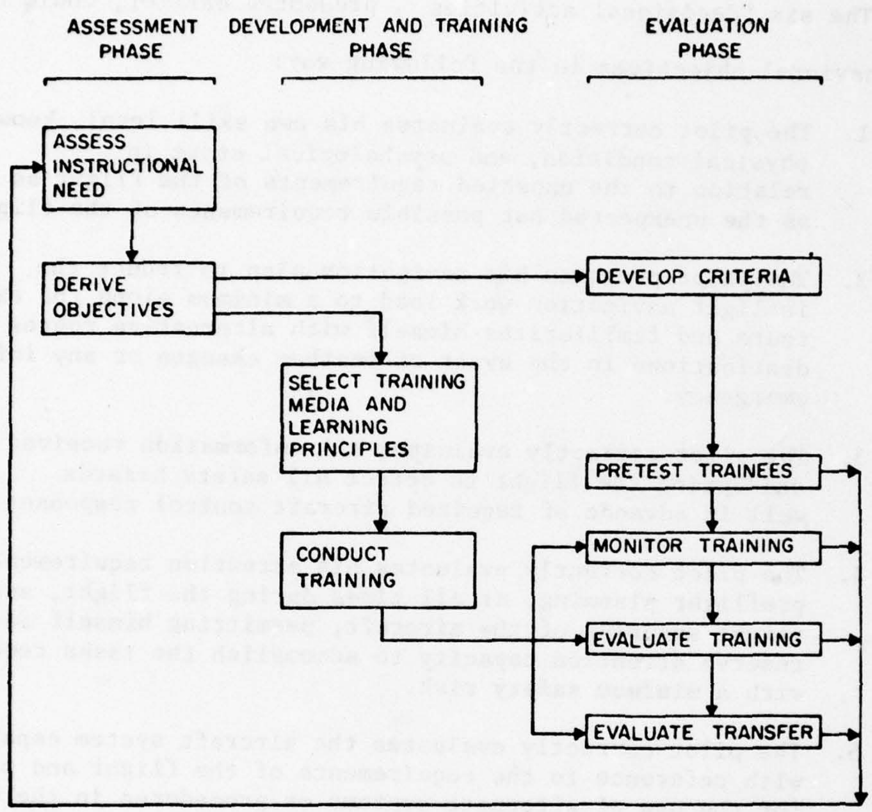


Figure 5. An instructional system.

assessment of the need for pilot judgmental instruction first requires an analysis of the present and future requirements for pilot judgment in the National Aviation System. Second, this assessment requires an analysis of the judgmental task from the behavioral standpoint, and third, an analysis of human attributes necessary to perform the judgmental tasks. These three analyses provide the bases for the development of judgment behavioral objectives.

The six "decisional activities", presented earlier, could be stated as behavioral objectives in the following way:

1. The pilot correctly evaluates his own skill level, knowledge, physical condition, and psychological state in relation to the expected requirements of the flight as well as the unexpected but possible requirements of the flight.

2. The pilot prepares his navigation plan to reduce the inflight navigation work load to a minimum along the expected route and familiarizes himself with alternative routes and destinations in the event of weather changes or any inflight emergency.

3. The pilot correctly evaluates all information received before, and during the flight to detect all safety hazards well in advance of required aircraft control responses.

4. The pilot correctly evaluates his attention requirements during preflight planning, at all times during the flight, and in post flight securing of the aircraft, permitting himself adequate reserve attention capacity to accomplish the tasks required with a minimum safety risk.

5. The pilot correctly evaluates the aircraft system capabilities with reference to the requirements of the flight and provides for the use of alternate systems or procedures in the event of failure in any component of the aircraft system.

6. After making the five evaluations and/or preparations mentioned above, the pilot adjusts his flight mission priorities to assure that the risk incurred is consistent with societal norms.

Although this list is preliminary and certainly incomplete, it follows the general guidelines for behavioral objectives given by Goldstein. In general, the behavioral objectives should specify what the trainee will be able to accomplish when he successfully completes the instructional program. They should also indicate the conditions under which the performance must be maintained and the standards by which the trainee will be evaluated. Thus, they provide direct inputs both into the evaluation phase and the development and training phase of the instructional model.

The Development and Training Phase. The development and training phase consists of the establishment of a training program (to achieve the behavioral objectives) and the conduct of the training. The development of a judgmental training program requires a blend of learning principles and media selection based on the skills, concepts, and attitudes that are to be transferred to the operational flying environment. A quotation from Gilbert (1960) is particularly appropriate in this regard:

> "If you don't have a gadget called a teaching machine, don't get one. Don't buy one; don't borrow one; don't steal one. If you have such a gadget, get rid of it. Don't give it away, for someone else might use it. This is a most practiced rule, based on empirical facts from considerable observation. If you begin with a device of any kind, you will try to develop the teaching program to fit that device (p. 478)."

Obviously Gilbert's remarks are quite extreme considering our previously established cost constraints. However, because of the nature of the subject matter involved in pilot judgmental training, a great deal of thought and perhaps empirical investigation should precede the design of the training environment for this type of instruction. We will see later that many of the basic principles of learning, on which most "teaching machines" are based, e.g., practice and reinforcement, may not apply in the usual way to pilot judgmental training.

The final step in the development and training phase is the actual conduct of the training. However, training is not an end in itself. The trainees and the training program must be continually monitored and evaluated. As shown in Figure 5, evaluation data are fed back to the assessment phase for use in the formulation of training program modifications responsive to the changing needs of the trainees and the environment.

The Evaluation Phase. The evaluation phase consists of the systematic measurement of changes brought about by the training program. The evaluation of a judgmental training program requires an assessment of the amount of measurable change in the trainee's judgment between the beginning and the end of the program. Thus, the evaluation phase requires the establishment of measures of success (criteria), based on the behavioral objectives, and the measurement of judgmental behavior both before and after the training process. Although rarely seen, data from these evaluations are vital to the success of any training program. As Goldstein (p. 23) points out, "...instructional programs are research efforts that must be massaged and treated until the required results are produced."

In pilot judgmental evaluation there also is a requirement for the assessment of judgment capabilities and tendencies relative to what society expects of pilots. This assessment is an absolute evaluation to determine whether or not a person has judgmental capabilities sufficient to serve as an aircraft pilot with passengers on board. It would be included as a part of the pilot certification process. Such an evaluation requires a knowledge of societal demands concerning pilot judgmental capabilities, at least qualitative criteria against which to judge the candidate, and an unbiased observation of performance to determine whether or not the candidate meets the criteria.

Some Learning Principles

Because of the common misapplication of some well established learning principles in many training programs, a discussion of these principles

as applied to pilot judgmental training is needed. Perhaps the most popular of these is really an assumption and not a principle at all. It is the assumption that "the best way to learn a performance is to practice that performance" (Gagne, 1962). This assumption is rooted in much of the educational literature and is often identified by the catch-phrase "learning by doing." Gagne points out that it may also be a generalization of the research on the conditioned response in which learning, particularly in animals, appears to have occurred only after a response (practice) has been made.

However, Gagne argues that practice is not an effective training method even for the acquisition of such motor skills as field gunnery. He says that "instructions about the correct sighting picture for ranging is much more effective in bringing about improved performance" than is practice on the gunnery task. The point is that training should emphasize the principles and procedures (or thought processes) involved, and practice should be directed to take advantage of these principles or take a minor role. If this is the proper emphasis for teaching motor skills, it is even more important in teaching judgmental skills based on thought processes (called "headwork," "thinking ahead," or "staying ahead of the airplane" by flight instructors).

A second learning principle that is frequently misapplied in training situations is variously called reinforcement, feedback, or knowledge of results during practice. This principle has been found to be most effective in choice behavior. However, Gagne points out that some

manipulations that artifically improve feedback during practice failed to show reliably better transfer to the operational environment, and others showed negative transfer. Apparently the form of the feedback is important.

As any beginning flight student will tell you, the usual feedback information such as "you did it right" or "you did it wrong" is almost useless. The time period between trials may be long, it is often cluttered with interfering information (Adams, 1976), and the trials themselves are often so complex that the student learns very little from such a response by his instructor. The student really needs to know <u>why</u> he did it right or wrong. He needs to know what rules he should have followed and where he strayed from those rules. This is important for perceptual-motor training, but it appears to be even more important for pilot judgmental training.

Gagne (1962) suggests that the following psychological principles are useful in training:

1. Any human task may be analyzed into a set of component tasks which are quite distinct from each other....

2. These task components are mediators of the final task performance; that is, their presence insures positive transfer to a final performance, and their absence reduces such transfer to near zero.

3. The basic principles of training design consist of: (a) identifying the component tasks of a final performance; (b) insuring that each of these component tasks is fully achieved; and (c) arranging the total learning situation in a sequence which will insure optimal mediational effects from one component to another.

These principles are guidelines for use in formulating the type of feedback that should be offered in pilot training. Such things as task analysis, intra- and inter-task transfer, and component task achievement

are needed to provide an efficient and thorough instructional process. Although practice and right/wrong types of feedback may be useful in some training situations, they should be de-emphasized in favor of these "thought" oriented teaching principles in all types of pilot training, but especially in judgmental training.

Transfer of Training

The essence of training is the impartation of skills, rules, concepts, or attitudes that transfer from the training environment to the operational environment. Therefore, the great emphasis in training research and training program design is concentrated in the area called transfer of training. A training medium or method may have positive transfer if trainees perform better than control subjects, negative transfer if trainees perform worse than control subjects, or zero transfer if trainees perform no differently from control subjects in the operational or transfer task. Transfer effectiveness of a training program is best measured using the transfer effectiveness ratio (Roscoe, 1971; 1972) which is quantitatively expressed by:

$$TER = \frac{Y_o - Y_x}{X}$$

Where:

Y_o = time to criterion in the transfer task for the control group,

Y_x = time to criterion in the transfer task for the experimental group,

X = time spent in training by the experimental group.

There are two proposed theoretical bases for transfer of training: identical elements and transfer through principles (Goldstein, 1974). The theory of identical elements was originally proposed by Thorndike and Woodworth (1901). They suggested that transfer occurs because of identical elements, e.g., aims, methods, and approaches (later simplified as stimuli and responses), in the training and the operational situations. Thus, the most positive transfer would occur when most of the task elements in the training situation matched those of the operational situation. This theory is often used, somewhat erroneously, as the basis for the design and use of high-fidelity pilot training simulators.

However, the theory of identical elements does not account for all that is needed in transfer of training. Some have argued that transfer can occur in situations other than those having identical elements. In these situations general principles, procedures, attitudes, and even general abilities are taught. Hendrikson and Schroeder (1941) report an experiment using a transfer-of-principles procedure relating to the refraction of light. Two groups of subjects received sufficient shooting practice to hit an underwater target consistently. Then the depth of the target was changed. One group then was taught the principles of refraction of light. In the next session this group performed reliably better than the group not taught the principles. Thus, the basic concern in the design of training environments should be which best helps the trainee to learn appropriate principles for application in transfer situations.

Judgmental Training Techniques

The point has been made repeatedly that, because of the nature of the subject matter to be taught (i.e., attitudes, principles, and motivations), the primary load of pilot judgmental training must be borne by the flight instructor. Practice and conventional self-teaching techniques (e.g., solo flying in the practice area) are highly inefficient methods for imparting these concepts. Webster C. Todd, Chairman of the NTSB, (Aviation Consumer, 1977) has said that much of flight training today is like "training people to jump through hoops. . . . you describe the maneuvers . . . and do those maneuvers. There's no attempt to get at judgment. You don't teach people judgment by teaching them to jump through hoops."

Given the emphasis on transfer of *principles* in judgmental training, an important question that must be addressed is whether current flight instructors are properly equipped to teach judgment? Todd's answer is no, at least, in most cases. The reason, he says, is that ". . . 99.9 percent of emphasis is on the first word (flight). There is no emphasis on whether that guy can communicate, whether he has the ability to transfer his skill or his knowledge."

Thus, the challenge is two-fold: instructors must be trained and motivated to teach judgment, and devices or techniques must be developed that permit adequate judgmental instruction with less than perfect flight instructors. The following is a discussion of some judgmental training techniques that could be applicable to pilot training in civil aviation. Most of these are viewed from the perspective of present programs such as

ground school, flight simulation, and flight training. Others, such as computer-assisted instruction, are relatively new to aviation.

Ground School. There are a number of excellent and proven ways that pilot judgment could be taught from the perspective of the conventional ground school. To afford it proper emphasis, it is suggested that judgment should be given a special section of ground school with the same status as meteorology, navigation, and Federal Air Regulations. This section could include lectures and/or discussions of aviation accident scenarios in which the pilot was a cause or factor, interactive movies, video tapes, slide presentations requiring student judgmental responses at critical points in flight scenarios, and independent study of the principles involved in good pilot judgment.

In addition, this ground school section could include instruction in information integration and subjective probability estimation of the type used in training people to be expert judges (Goldberg, 1968). Judgmental behavior in expert judges is characterized by a chunking or the formation of clusters of stimulus attributes and response alternatives for economy in the thought process. Ground school students could be taught to use these principles in their judgmental processes. For example, the instructor would show how various types of probabilistic information (stimuli) such as weather forecasts, predicted aircraft system malfunctions, and predicted Air Traffic Control problems should be combined.

These data would be used to establish alternatives (responses) that would be evaluated to form a decision at any time before or during a flight. The instructor would teach the student how to "think ahead" or anticipate

decisions that might have to be made later resulting from present choices of action. Such anticipation permits the gathering of relevant information under lower levels of stress, when errors are less frequent than later in the flight when time to decide becomes a constraint (Janis and Mann, 1976).

Finally, this section of ground school could include decision-making training of the type suggested by Janis and Mann (1977). Their approach is based on a "conflict-theory" model of decision-making. This model was used to develop methods of fostering improved decision-making skills, and these methods were subsequently tested and found to be successful. Figure 6 (from Janis and Mann, 1977) presents a summary of the model and the way in which various conditions lead to good decision-making patterns (Unconflicted Adherence, Unconflicted Change, Defensive Avoidance, or Hypervigilance).

In a state of "Vigilance" the decision maker searches for relevant information, assimilates it in an unbiased manner, and appraises alternatives before reaching a decision. He plans for all foreseeable contingencies. Decision makers exhibiting the patterns of "Unconflicted Adherence" or "Unconflicted Change" are manifesting boredom or complacency and uncritically continue what they are doing or change to a new course of action without examining and appraising all available information or planning for contingencies.

While these two patterns do not usually result in disasterous consequences, they can lead to defective judgment. When exhibiting the "Defensive Avoidance" pattern, the person tends to avoid making a decision by procrastinating, shifting the responsibility to someone else, or by

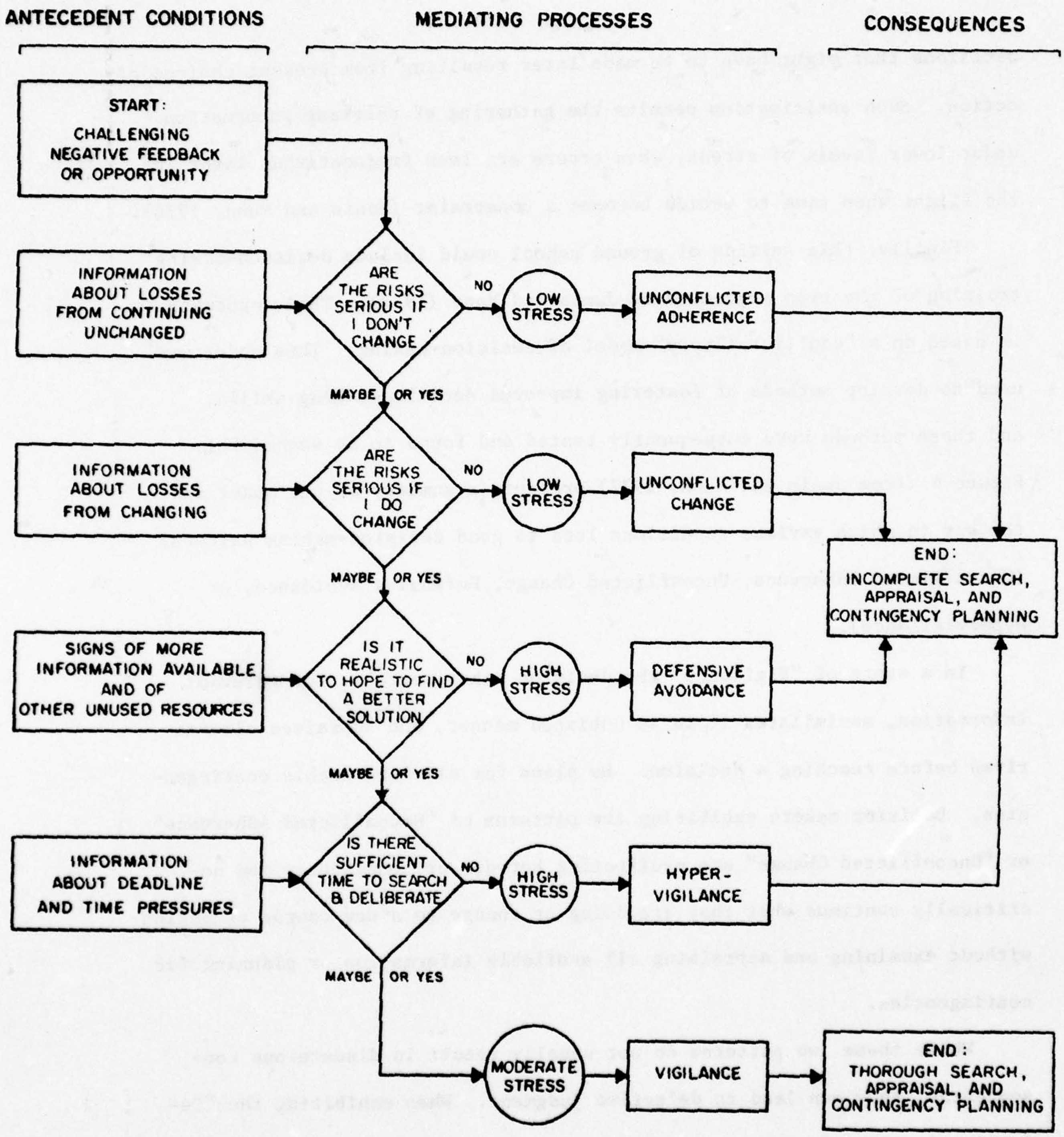


Figure 6. Conflict-theory model of decision-making (Janis and Mann, 1977).

wishful rationalizing. Finally, the "Hypervigilance" pattern is marked by emotional excitement, cognitive restriction, and adherence to simplistic solutions. Extreme hypervigilance is synonymous with panic.

The authors provide evidence that the "Vigilance" pattern of judgment provides the most adaptive decision processes due to the facts that 1) all available information and alternatives are carefully appraised, 2) there is thorough planning for contingencies, and 3) even after making the decision, the person is open to new information. Thus, judgmental training should concentrate on eliminating or minimizing the factors resulting in the four less adaptive decision patterns.

Such training may be accomplished by teaching the decision maker to: (1) evaluate realistically the risks involved in a given course of action, (2) become aware of all possible alternative courses of action, (3) provide sufficient time to make a judicious decision, and (4) become aware of faulty patterns of reasoning and rationalizations that may degrade his judgment. Experimental evidence is cited to indicate the efficacy of this approach in fostering vigilance and, therefore, improving the quality of judgment.

Two of their suggested approaches to improving the quality of judgment are "emotional role-playing" and the "balance sheet" procedure. Emotional role-playing is a psychodramatic technique in which the student plays the role of the pilot who makes the wrong initial decision. As the scenario continues, the student becomes aware of risky circumstances and general loss of control of the situation that result from making the wrong initial decision.

Janis and Mann (1976) report that this technique is capable of producing long-term changes in attitudes concerning personal vulnerability to low probability consequences (lung cancer) in the case of cigarette consumption among heavy smokers. These results are similar to those reported by Rubinsky and Smith (1972) who found that subjects who experienced a simulated accident as part of their training program were less likely to repeat the hazardous behavior in the operational setting. In the context of the ground school these techniques would probably be best carried out through interactive films or computer-aided instruction programs.

A second approach used to bring about the vigilant form of decision behavior, the "balance sheet" procedure, is a "predecisional exercise that requires a decision maker to confront and answer questions about potential risks and gains he had not previously contemplated" (Janis and Mann, 1976). Four general areas are suggested for consideration using balance sheet techniques. In pilot judgmental training these four areas could be monetary gains or losses, personal or passenger safety and comfort, self-approval or disapproval, and approval or disapproval of others.

Gains or losses due to alternative courses of action are listed on the balance sheet for each of these four areas. Such an exercise in a ground school situation would help the student to prevent errors of omission, to develop correct mental patterns of information search and evaluation, and to avoid the use of certain types of information (i.e., approval or disapproval offered by self or others) which, at best, are irrelevant and at worst could lead to an unsafe flight decision.

Computer-Assisted Instruction. An instructional technique that holds unusual promise for pilot judgmental training and evaluation is computer-assisted instruction (CAI). The great disadvantage and a source of the resistance to implementation of CAI devices in pilot training has been their initial cost and limited availability. However, recent technological advances are making them available for widespread use at a relatively low cost (Trollip and Roscoe, 1972; Trollip and Ortony, 1977; Finnegan, 1977).

Although CAI programs and devices are available in several forms, the dialogue systems that permit student-computer interaction unrestricted by preset response alternatives, show the greatest potential for application to pilot judgmental instruction (Alpert and Bitzer, 1970). These systems depend on a set of stored algorithms that are used by the computer to construct unique responses to student questions. In addition, student responses are not limited to exact duplicates of prestored expected responses. The program understands a variety of student responses and is able to proceed accordingly.

Cooley and Glaser (1969, pp. 574-575) give a series of steps taken in most CAI programs that closely parallel the instructional model presented in Figure 5, (page 47).

1. The goals of learning are specified in terms of observable student behavior and the conditions under which this behavior is to be manifested.

2. When the learner begins a particular course of instruction, his initial capabilities -- those relevant to the forthcoming instruction -- are assessed.

3. Educational alternatives suited to the student's initial capabilities are presented to him. The student selects or is assigned one of these alternatives.

4. The student's performance is monitored and continuously assessed as he learns.

5. Instruction proceeds as a function of the relationship between measures of student performance, available instructional alternatives, and criteria of competence.

6. As instruction proceeds, data are generated for monitoring and improving the instructional system.

Although practice and feedback are frequently used concepts in CAI programs, these could be augmented by presenting principles, rules, and reasons for taking certain courses of action. In a pilot judgmental training program the student could be presented with a flight situation requiring judgment. The student could be asked to respond by listing all of his alternatives and the factors affecting each. He could even be asked to estimate his probabilities of success for each alternative. The computer could then examine the flight experience data on the student (which he had entered previously) and the stored accident statistics from similar circumstances for recent years in U.S. Civil Aviation. The computer would then respond with comments on the appropriateness of the student's responses, the alternatives that may have been omitted, and the principles that should have been followed in making the correct decision. The program could then branch to another problem, the difficulty of which would be based on the level of judgmental capability evidenced by the student's responses to the previous problem.

Complexity and realism could be added to the judgmental task by the addition of a simple flight hand-controller and an airplane symbol with a

map on the screen. The purpose of this controller would be to provide indications of progress toward a destination and time available for decision, not for instruction in flight control. In one experiment, Trollip (1977) has demonstrated the utility of teaching the cognitive aspects of flying a holding pattern in various wind conditions using such a device. In a subsequent experiment, Finnegan (1977) demonstrated that such training transfers reliably to flight.

CAI has many advantages not commonly associated with other instructional techniques. The most important of these is individualization of instruction. It can adapt to the specific needs of the individual and interact with the student at his current level of ability (Goldstein, 1974). Second, the unencumbered reinforcement capabilities of CAI are a real benefit to the student. It has no personality or ulterior motives to clash with those of the student. Third, CAI systems do not require the presence of a teacher, although it may be beneficial to have one close by for consultation. Fourth, it permits standardization of instruction across a wide area. One central computer could potentially support terminals at every pilot instruction center in the United States at a relatively low cost (on the order of $2.50 per hour). No student would be handicapped by a bad instructor who underscores weaknesses in the simulator. Fifth, data gathered from student responses could be stored as long as necessary for use in updating instructional programs or in evaluating pilot judgmental capabilities. Hard copy outputs of the data could be made available with a command from any terminal.

CAI also has a number of limitations. First, very little is known about the effectiveness of the instructional techniques described above. Research is still needed to program such an instructional system most effectively. Second, large outlays of money would be required for hardware to begin such a program. This cost might be eased by making it a part of the emerging automated pilot briefing system. Third, some users might object to the requirement for communication with the computer via a keyboard, although keyboards rapidly are becoming a part of the pilot's way of life. Fourth, Goldstein (1974) expresses concern for the effects of a machine-oriented learning environment on satisfaction, motivation, and development. Furthermore, it is possible that some initial consultative help would be required while students were learning to use CAI.

Flight Simulation. Of the various alternatives available for pilot training, it is becoming apparent that procedural, perceptual-motor, and judgmental instruction in ground-based aircraft simulators may be the most viable for many operators in civil aviation. The education of pilots for military and airline applications has become increasingly dependent upon ground-based aircraft simulators. As stated by Williges, Roscoe, and Williges (1973, p. 5):

> Synthetic flight training has come a long way from the Barany chairs and stub-winged Jennies of World War I. Although never submitted to formal, objective experimental evaluation during their heydays of World War II, there is now ample evidence that the Link C-3 and AN-T-18 "blue boxes" and their close descendants, the School Link, 1-CA-1, 1-CA-2 (Navy, SNJ; Air Force, P-1), the Frasca, and the GAT-1, were and are highly effective training devices in the hands of an imaginative flight instructor.

> Modern flight simulators, devices that represent specific counterpart aircraft, also unquestionably have high potential transfer effectiveness, although all studies to date have tended toward demonstration rather than controlled measurement of transfer. In particular, an 89 percent reduction in helicopter flight time followed the introduction of a new training curriculum using the 2-B-24 synthetic flight training system (SFTS) in the U.S. Army's undergraduate instrument pilot training program (Caro, 1973). Forty-three hours of training in the 2-B-24 device replaced 53.5 hours in the TH-13-T and UH-1-H helicopters, for a cumulative transfer effectiveness ratio of 1.24.

Despite statistics showing them to be highly effective training devices for military and airline applications, simulators in beginning pilot training programs of general aviation have been less effective and less popular. A number of flight schools use simulators in their pilot training programs with varying degrees of success, but the majority of the approved flight schools do not use them at all. Many of those schools that do use simulators have not approached the transfer effectiveness level predicted in the research literature.

There are a number of reasons for the relatively low popularity and poor showing of simulators in general aviation pilot training programs. These include initial costs, space requirements, maintenance costs, and unreliability. But perhaps the primary reason is the lack of instructor ability and motivation to use the simulator in disciplined, innovative flight training of the type described earlier in this report. Flight instructors typically aspire to higher flying positions that can only be attained through hours spent in the air. Therefore, they are not motivated to provide inspiring simulator instruction. The instructor's attitude toward simulator instruction may be transferred to the student, and as a result the instruction is ineffective.

Even if the flight instructor is motivated to give good instruction, he may not be aware of the most effective ways to teach flying using a simulator. Very few schools that use simulators specify how they are to be used. Typically, instructors are told to spend a certain amount of time in the simulator with each student, but instructional methods and procedures are left to the individual instructor to create. Consequently, the simulator may be used only as a substitute device for practicing aircraft control. This use of a simulator is wrong on two counts. First, as noted earlier, unstructured practice is a poor learning technique. Second, learning how to fly the simulator has no redeeming value -- it is the airplane that one must learn to fly.

Using the simulator as a practice device also leads to low standardization of instruction and the placing of greater emphasis on weaknesses in the simulation than on specific flight training objectives to be accomplished in the simulator. A good instructor with a blackboard, chalk, and model airplane could probably show better transfer effectiveness than a poor instructor with a high-fidelity flight simulator. Good simulator instruction, whether the subject matter is slow flight or pilot judgment, requires the same creativity and thought that one would use in teaching flying principles with a blackboard and chalk.

It is worth noting at this point that, by nearly all educational principles, an airplane cockpit is a particularly poor learning environment due to distractions

from several different sources. The airplane cockpit environment usually has a high auditory noise level permitting communication only by shouting, movement is restricted by space limitations and seat belts, the student's anxiety level may be high, radio communications interrupt cockpit communication, turbulence and wind conditions may make certain maneuvers difficult to teach, and visual distractions from the need for traffic avoidance continually interrupt the teaching process. Finally, the variety of flying environments is severely limited in the airplane. For example, in some parts of the country there is inadequate exposure to a complex, high density traffic environment, a type of training that is becoming increasingly important.

Undoubtedly, some instruction in the aircraft is necessary in primary flight training, but one would think that education under these circumstances should be minimized in favor of the much more safe, dependable, energy efficient, and civil environment of the flight simulator and other ground teaching aids, especially for teaching cognitive items such as judgment. However, except for their highly effective use in airline training, simulators have been virtually unused in the teaching of decisional skills. Furthermore, there has been very little research on the question of whether or not a simulator could be an effective judgmental training aid.

Although judgmental training could be interspersed throughout, it is an important topic that deserves the status of a special section of the simulator training program. In some ways judgmental training in a simulator would be expected to be more cumbersome than in ground school or

CAI because, at least in current practice, it depends on the instructor to create the simulated flight situation primarily through verbal communication. Nevertheless, the simulated flight environment provides an additional opportunity to teach judgmental principles, if properly structured, in a somewhat more realistic environment than ground school and CAI can provide.

Probably the best way to begin judgmental training in the simulator is to use the airline approach, teaching procedures that are to be followed in each situation that departs from normal flight. This includes system failure detection as well as establishing courses of action to correct or counter system failures. Principles involved as well as corrective procedures would be taught according to this method, and appropriate judgmental performance measures could be developed.

The simulator instruction could also include the creation, by the flight instructor, of judgment-demanding situations that do not involve the failure of systems. These situations would demand decisions about whether or not to continue flight into deteriorating weather, decisions about passenger demands for landing at an unfamiliar alternate airport, decisions about weight and balance considering field conditions, density altitude, etc. In all cases the instructor would ask the student to state several alternatives available to him and also to state which he would choose. These situations could be developed from NTSB accident briefs, and they could be a part of the flight instructor's simulator judgmental instruction package.

In conclusion, the simulator judgmental instruction should be treated as a serious and vital part of the flight student's training. The simulator

must be treated as an important training aid just as the airplane and the blackboard are treated. The instructor has the opportunity and responsibility to instill serious, mature flight attitudes in his students by his approach to judgmental instruction in the simulator. The simulator provides an outstanding medium for teaching a student that it is not socially demeaning to turn around in the face of deteriorating circumstances. But the training is only as valuable as the instructor's approach to simulator instruction is serious.

The Airplane. Of all the media available, the airplane is probably the most difficult to use for direct, systematic judgmental training. The reason is that for the sake of safety, convenience, and cost most judgmental problems must be halted before the student sees the final consequences of his decisions. He frequently must take the instructor at his word that his decision would have resulted in a safe or unsafe situation. However, the airplane offers special opportunities for judgmental instruction because the environment is more realistic, it is more meaningful, and therefore, it is more likely to cause a more permanent behavioral change in the student than other training media.

Everything that has been said about instructor attitudes and approaches to judgmental instruction applies doubly when actually flying an airplane. Without a systematic judgmental training program, good pilot judgment is acquired by the cautious and the lucky over years of flying experience in many varied situations. Our task as flight instructors using systematic judgmental training techniques should be to compress a lifetime of flying experience into a relatively short training program to instill good pilot judgment into the emerging private pilot.

To accomplish this in the airplane requires a consistent, disciplined flight instructor who always follows the rules that the student is expected to follow, or he provides a good explanation for why he deviates from them. It also requires the instructor to adhere to the learning principles stated earlier, that practice and feedback are not highly beneficial unless they are accompanied by directions and explanations that are logically structured and clearly stated.

Judgmental instruction in the airplane frequently takes the form of simulated or imaginary situations requiring the use of judgment, and such activities should be interspersed throughout the flight training program. This is already being done to some extent through training in simulated engine failures, other system failures, and all types of stalls. This training could be expanded to include many of the hypothetical situations discussed above for use in the simulator. Portions of such simulated situations could be a part of every instructional flight.

For example, the instructor could tell the student that today our takeoff is from a 2000 foot runway with a density altitude of 6000 feet, a zero wind, and two extra passengers who must fly to Walla Walla. The student would be asked to state his alternatives and the probability of success for each alternative. Such exercises conducted seriously would instill in the student a sense of responsibility for the safety of the flight and better information integration procedures -- and better judgment.

It is the instructor's responsibility to teach the student that it is not socially demeaning to refuse to fly or to turn around in the face of deteriorating circumstances. Such situations should be made to occur

several times during the student pilot's instruction program in the airplane. Pilots have often said that it is most difficult to turn around the first time. After that, it becomes easier. In this regard, it is important to teach the student how to avoid the tremendous social pressure that a group of important passengers can exert. The pilot must be willing to isolate himself from his flight naive passengers in all important decisions.

Finally, often one of the most difficult evaluations a pilot has to make is the self-evaluation of his own skill, knowledge, and judgmental capability relative to a proposed flight. To assist him in this regard, the pilot should develop a list of general limitations on flight procedures based on his own capabilities. These limitations must be applicable to all flights regardless of who the passengers are. They should be invoked during a rational moment, and the pilot's resolution should be strong enough to withstand the enormous social pressure to deviate from them either before or during a specific flight.

Situational Emergency Training

The Air Force has begun a research program (Thorpe, Martin, Edwards, and Eddowes, 1976) aimed at improving pilot decisional processes during emergency situations. Although the goals of this program are more limited than those of judgment training in Civil Aviation, the approaches suggested for training are very similar to those we suggest for judgmental training. The proposed training program being studied called Situational Emergency Training (SET) has replaced the traditional Boldface procedures of other USAF weapons systems, which must be committed to memory, in the F-15 training program. Although Boldface procedures are effective in many situations where their solutions are applicable, the authors suggest that there are situations in which these solutions do not apply and such training inhibits good judgment in such situations.

SET, on the other hand, encourages the development of judgment and centers training around three emergency rules: (a) maintain aircraft control, (b) analyze the situation and take proper action, and (c) land as soon as practicable. The underlying concept of SET is situational training. The pilot is taught to discriminate the relevant from the irrelevant dimensions of situations which are systematically manipulated in the training program. As we have pointed out above, this discrimination process is fundamental to good judgment. The authors suggest a scenario development program using instructor training courses as one of the major sources of input. These scenarios would be used in the training program to manipulate dimensions of emergency situations presented to students.

As presented in our definition of judgment, these authors suggest that pilot judgment has two components, one relating to discrimination or the "process of sorting or attending to specific stimuli", and the other relating to "response development and selection". In answer to the question "What is it in a pilot's experience that develops judgment?", these authors answer:

> Part of the answer is that during the period of accumulating flying hours the pilot operates his aircraft in a wide variety of situations and in later flying he will again encounter the same or similar situations. Similarly, as he exchanges his own experiences with those of other pilots, he learns about new situations and how they were handled by others. Presumably, he also gains confidence that he can handle more complicated problems and this should aid him in resisting stress effects when confronted with an emergency situation.
>
> Situational training attempts to develop judgment by having the pilot exercise discrimination and response selection in training sessions. This surrogate flying experience permits the student to accrue judgmental skills by discriminating relevant and irrelevant dimensions of various emergency situations, and then to use the relevant information in the selection of a correct response. It is a systematic way of exposing the pilot, especially the less experienced pilot, to a wide variety of judgment building situations.

Outline for Judgment Training

The following is a summary of suggestions made for a three-phase judgment training program. One essential feature of this systematic approach is that the program is continually adjusted through the application of the results of the evaluation phase. Although the curriculum includes suggestions for use of four different judgment training media, this does not mean that all four are needed to improve pilot judgment. Furthermore, these suggestions are made on the basis of a review of related research literature and an opinion survey of aviation educators. Empirical data on alternative training techniques are needed before rule changes can be suggested.

I. Need assessment

  A. Analysis of training requirements

  B. Statement of desired behaviors

II. Development of training program

  A. Develop judgment training principles

    1. Thought oriented (attitudes, principles, motivation)

    2. Use of transfer of principles techniques

  B. Train instructors

  C. Develop curriculum

    1. Ground school

      a) Raise to a level of a section (as meteorology).

      b) Use media to introduce judgment situations.

      c) Sequence subject matter with subject matter being taught elsewhere (ground school courses, flight, etc.).

      d) Show specifically how to search for information, evaluate information, integrate information, establish alternatives, choose correct responses, in each situation.

e) Show effects of stress and other pressure factors on decision-making and show how to adjust for them.

2. Computer-assisted instruction

a) Use CAI to augment instruction in (1) above.

b) Sequence subject matter with subject matter covered elsewhere and student progress.

c) Use as an evaluator of judgment performance.

3. Simulator instruction

a) Begin with procedure training.

b) Use judgment principles in teaching all maneuvers.

c) Sequence unusual scenarios throughout training program using subject matter being taught at the time.

d) Include a special section for concentrated situational judgment training.

4. Aircraft instruction

a) Teach by example (follow rules or provide very good reasons for deviating from them).

b) Use judgment principles in introducing and teaching every maneuver.

c) Include a special section of flight training program for use of situational training of judgment.

d) Teach student how to evaluate and establish his own limitations during practice and to never exceed those limitations.

D. Conduct training

III. Evaluation

A. Pre-training evaluation

1. To provide an initial indication of judgment training requirements

2. Used as baseline in establishing the value of training program

B. Monitor training

   1. Evaluate progress at each introduction of judgment scenario

C. Evaluate training

   1. Test for progress in judgment performance

D. Evaluate transfer

   1. Test for value of judgment training in real-world environment

AD-A057 440   ILLINOIS UNIV AT URBANA-CHAMPAIGN SAVOY AVIATION RES--ETC   F/G 5/9

JUDGMENT EVALUATION AND INSTRUCTION IN CIVIL PILOT TRAINING.(U)

DEC 77   R S JENSEN, R A BENEL   DOT-FA77WA-3920

UNCLASSIFIED   FAA-RD-78-24   NL

2 OF 2
ADA
057440

END
DATE
FILMED
9 -78
DDC

PILOT JUDGMENTAL EVALUATION

The main objective of this section is to indicate whether and how pilot judgment, as defined, may be evaluated. The scope of the discussion will, in general, be limited by the model of an instructional system shown in Figure 5 (page 47).

A Definition

An evaluation may be defined as: to determine the worth of, to find the value of, to appraise. From this definition we may see that evaluating judgment inherently involves judgments of value. Sommer (1977) distinguishes evaluation from research, noting that evaluation makes use of research findings. Research often proceeds from hypotheses (e.g., Program X works well), while statements of hypotheses are studiously avoided during evaluation. It is critical to avoid any indication of bias prior to the evaluation. In further definition of the purpose of evaluation Sommer (1977) states, "The task is not to determine whether something is effective in the absolute sense but its effectiveness along different dimensions." This implies that evaluation may be made by individuals with varying interests according to those interests.

One of the major weaknesses in many evaluations of complex behavior is the frequent use of subjective judgments based on irregular and uncontrolled observations (Abeles, 1973). To avoid these problems several alternatives are available. When there is no objective measure by which an optimum evaluation can be determined, the optimum may be determined

by securing subjective evaluations from skilled individuals who have the particular competence and experience required. These judges will use a limited number of criteria to determine the optimum, but the relative importance or weights for these criteria may vary from one judge to another (Gulliksen in Shelley and Bryan, 1964). Abeles (1973) found that the achieving of a consensus among judges is difficult, particularly when they use impressions. To eliminate this source of variance, guidance should be provided to the evaluators in the particular weightings to use.

Can Pilot Judgment be Evaluated?

A cynic might state, "If it can be taught, it can be evaluated." However, there is very little evidence to suggest that pilot judgment is being systematically evaluated -- just as there is little evidence to suggest that it is being taught. Fortunately, there is evidence that complex judgmental behavior can and is being evaluated with varying degrees of success in other contexts. For example, McGuire and Babbott (1967) and Levine and McGuire (1968; 1970) simulated clinical problems that a physician might face and provided carefully structured environments in which to evaluate the physician's performance.

McGuire and Babbott's technique required initial information from a patient, involved sequential interdependent decision, allowed physicians to acquire relevant additional information, forced physicians to abide by decisions, and was adaptive to allow for varying decisions and outcomes of the decisions. Scoring was done by a group of experts classifying the judgments into five categories from poor to excellent. They also provided

a summary of responses including a measure of efficiency and both errors of commission and omission. They report that evaluation of clinical judgments can be made using this techinque.

Levine and McGuire (1968) demonstrated that judgmental evaluations may be made on the basis of role playing by candidates both before a medical board and before medical examiners. They concluded that this technique is more effective and lower in cost than direct observations. In a related study, Levine and McGuire (1970), using similar groups, found that the structuring of oral exams, the standardizing of case materials, and the training of examiners minimizes the problems of evaluation.

Although at first glance clinical judgments of physicians and in-flight judgments of pilots appear quite different, the underlying demands are reasonably similar. Both have the potential to be life or death situations, both impose enormous information processing demands on the judge, and both require rapid, sequential, interdependent judgments. It seems plausible to assume that techniques effective for evaluating one should be effective for evaluating the other.

In further support of the feasibility of pilot judgmental evaluation, the adaptation of signal detection theory (TSD) to the flight situation has been proposed. An individual's bias or threshold for response is analogous to the motivative aspect of pilot judgment. In structured situations, this bias before and after training should be measured to determine the effectiveness of the training program.

Finally, studies reported earlier in which subjects were taught to be expert intellective judges (Shanteau, 1975) and studies in which subjects were taught "vigilant" decisional responses (Janis and Mann, 1977) both present successful judgmental evaluation techniques. These evaluation techniques involve the comparison of observed responses in judgmental situations against the expected or correct responses in those situations. These same techniques could be used in pilot judgmental situations provided that proper behavioral responses or evaluation criteria are established.

Pilot Judgment Evaluation Criteria

As indicated by the instructional model, Figure 5, the most important initial step in the establishment of any evaluation procedure is the development of criteria against which the student's behavior can be compared. This is particularly true when behavioral evaluations are made subjectively and yet standardized over a national population of evaluators such as flight instructors and examiners. Thus, pilot judgment evaluation criteria must be meaningful, they must be objective, and they must mean the same things to a wide variety of evaluators, and to those being evaluated so that they may know what is expected of them.

As suggested by the instructional model, evaluation criteria, in training situations, should be derived from previously established behavioral objectives. Our list of six preliminary judgmental behavioral objectives, given earlier (page 48), are positive statements of acceptable pilot judgmental behavior. These were developed through an examination and integration of NTSB reports of accidents in which the pilot's decisional behavior was found to be

at fault. Although these NTSB reports are stated in the negative, they provide a useful basis upon which to build pilot judgmental evaluation criteria.

A few of the pilot defections reported by the NTSB that we classify as "decisional activities" are:

1. Operating an aircraft with known deficiences
2. Operating beyond own experience or ability qualifications
3. Continuing into known adverse weather under visual flight rules (VFR)
4. Selecting flight route over unsuitable terrain
5. Neglecting adequate preflight planning or preparation
6. Flying an unfamiliar aircraft
7. Allowing attention to wander from operation of aircraft.

Each of these frequently reported causal factors in accidents represents a qualitative evaluation of a pilot's judgment by the NTSB that could be used as the basis for a set of quantitative criteria in which the pilot's judgment is evaluated on a scale from good to bad, possibly quantitatively.

In pilot training, for each level of pilot experience, certain judgmental hurdles (proficiency levels) could be objectively specified. The instructor or examiner who evaluates the judgments would have a range of acceptable performances, also objectively specified. Evaluation of pilot judgment would be a matter of comparing performance against the established criteria. The critical point for judgment evaluation in a national system, such as that within which pilots operate, is the use of the same criteria by all judges as well as by the pilots themselves.

While understating the difficulty of constructing these criteria, an approach to this level of standardization is essential to the evaluation of training and transfer. For this purpose, NTSB reports may be used as one basis for judgmental evaluation criteria. As suggested in our definition, the ultimate standard against which pilot judgment is measured is determined by societal norms. Although one might argue with specific findings, the NTSB is officially sanctioned to assess the cause of accidents including those involving faulty pilot judgment. To make any such an assessment requires an assumption of what society would have done in that situation for use as the index of desired performance. Therefore, the NTSB findings provide a unique standard for the establishment of pilot judgmental evaluation criteria.

The Psychology of Evaluation

The history of formal psychological testing is as old as the history of formal psychology. The roots of this testing extend at least as far as the mid 19th century with the introduction of psychometrics by Sir Francis Galton. The roots of informal psychological evaluation are lost in antiquity and probably began as an answer to "I wonder why Torg did that?" Both forms of evaluation are still being practiced today. The formal evaluation procedures include questionnaires, interviews, objective tests of intelligence, personality, etc., projective tests (e.g., ink blot test) primarily for personality testing, and situational or job sample evaluations. Informal evaluation includes all the less well specified occasions during which we "size someone up."

The two conditions essential for any evaluation method are reliabilits and validity. Relability always refers to consistency or repeatability,

while validity indicates how well the evaluation measures what it is supposed to measure. One type of reliability is scorer reliability, or the measure of interscorer differences, often referred to as observer-observer reliability. Stated another way, do two people see it the same way? Efforts must be made to ensure that evaluations made by different evaluators are very nearly equivalent. Another kind of reliability, or more accurately, another source of unreliability, refers to the consistency or inconsistency with which the same pilot performs in the same situation on different occasions, often called ride-ride reliablity.

Validity must be determined with reference to the particular use for which the evaluation method is being considered (Anastasi, 1968). Content validity involves a systematic examination of the test content to determine whether it covers a representative sample of the behavior domain to be measured. This is built into the test and is primarily dependent on the development of appropriate criteria. As an example, the content of a flight check during primary pilot training should sample the content of the previous instruction. Not only should a proper proportion of items be included on such a check, but irrelevant items should be eliminated.

Criterion-related validity, indicates the ability of the test to make predictions of behavior in future situations (predictive validity). Tests may be very reliable but useless without validity. As an example, we might weigh a student pilot every day and find his weight remains essentially constant. Our testing instrument (the scale) is consistently accurate, but it is unlikely to yield a valid measure of a pilot's performance. On the

other hand, an unreliable score can never be relied upon as an index of validity.

Evaluation might be considered as the end product of a systematic attempt to observe human behavior. However, it is necessary to exert a certain degree of selectivity of observation. It would be both impractical and inefficient to follow an individual day after day recording everything he does. The individual might not spontaneously exhibit the types of behavior in which we have interest, and we would rarely observe the same behavior in other individuals for comparison. One way to control this problem is to provide individuals with one or more specific tasks to perform. The nature of the tasks may be varied to elicit the behavior we wish to observe.

An individual task of this sort is often viewed as the test item and a series of these tasks as a test. To trust results from a test we need measures of its accuracy and utility (reliability and validity). Generally, the more behavior samples taken, the more accurate the measurement will be (Lord and Novick, 1974). Accuracy is critical when we use measurements to discriminate among test-takers. However, we must weigh the increase in accuracy attainable against the increased cost of taking more samples of behavior. The utility is generally a function of careful selection of observations. Careful selection of our behavior samples will increase the utility or validity of our measurement.

A number of methods have been used by psychologists, educators, and personnel managers to evaluate human behavior and personality. These methods include interviews, questionnaires, written tests, and situational

tests with objective response criteria. Each of these may have some utility in pilot judgmental evaluation prior to the beginning of flight training, during judgmental training, and in the measurement of transfer. Those methods useful in initial testing would be particularly valuable in the identification of potential judgmental weaknesses in beginning students correctable through various instructional techniques.

The Interview and Questionnaire. Although the structured interview has limited applicability in judgmental evaluation because it requires trained interviewers, it should be mentioned because of its widespread use in other contexts (Bellows, 1963) and because it may be useful in some pilot training situations. In the formal sense interviews are most commonly used in personnel selection or psychiatric screening, but we all use many of the same techniques informally. Much of the face to face interaction between instructor and student pilot could be evaluated if transformed into what would amount to responses to a formally structured questionnaire.

Common errors of the interview are likely to occur in both formal and informal assessments. A personal bias of the evaluator may cloud his evaluation, and he may base his evaluation on irrelevant personal characteristics of the individual rather than his statements. The way in which the situation is structured may have a profound effect on the evaluation of the individual's judgment. The individual being evaluated may not recognize the relevant dimensions of the situation because of the format. Despite being highly motivated to render an adequate judgment, he might be unable to do so in the typical interview situation.

Another common problem in interviews is the establishment of rapport between interviewer and interviewee. While the instructor and student pilot may have sufficient time to develop rapport, it is conceivable that changes in evaluators will not retain this established rapport. Any deleterious effects will, likely, be magnified when the situation becomes as important as a flight check. Small change in the evaluator's approach may produce dramatic changes in performance which would be reflected by changes in evaluations.

In addition to these problems, there is a wealth of data that calls into question the evaluations of highly skilled interviewers. In certain cases interviewers may escape common pitfalls of interviewing precisely because they don't know enough about the criteria to be biased. The value of the interview as an evaluation method is extremely limited (it may even be negative), but there are justifications for retaining face to face contact between evaluee and evaluation. Strauss and Sayles (1967) indicate that interviews, as performance evaluations, can be considered as a form of coaching and may be useful as part of the overall training program. Interviews are also good because they provide feedback from the evaluee to the evaluator. Evaluees should be encouraged to point out any weaknesses they perceive in the interviewing process.

Objective Tests. Objective tests may be used to measure many diverse entities. The type of test that most people think of first is likely to be an intelligence or I.Q. test. Anastasi (1968) reviews objective tests and notes that the history of intelligence testing predates Binet by at least 21 years, beginning in 1887. Standardized aptitude and achievement tests are common in our society from early school years through the search for employment.

The typical I.Q. test measures an individual's performance on a wide variety of questions or tasks and normally provides one score called the Intelligence Quotient, or I.Q. Part of what we have defined as judgment resembles part of what is known as intelligence. To this extent, I.Q. tests may give information about the ability to make judgments. As an example of this limited correspondence, an item on the Wechsler-Bellevue (now updated as the Wechsler Adult Intelligence Scale) asks subjects what they would do if they found an addressed, stamped letter in the street (Wechsler, 1939). Responses to this and similar items are scored for the comprehension subscale, but their relationship to judgment, as good sense, appears evident.

The objective personality tests or self-report inventories number in the hundreds. These tests purport to measure emotional, motivational, interpersonal, and attitudinal characteristics. Shealy (1973) and Johnson and Davis (1972) have compared risk-taking behavior with several standard tests of this type and have found them to have some value. However, the work on this type of test has not generally been aimed at pilot judgment. Borg (1963) indicates a difficult problem with some self-report inventories. Using these tests may result in subjects' faking replies to create the desired impression. Fortunately, many of the standardized tests are constructed to catch faking.

Situational Testing. All testing involves the examination of a sample of behavior. A logical direction to take would be to consider behavior which resembles that exhibited in the criterion situation. This approach enters the field commonly referred to as situational testing or job sampling. More generally, we are referring to simulations with varying degrees of fidelity from the precise duplication of real-life flight

tasks to the use of mockups that are analogous to but do not greatly resemble the actual situation.

Evaluating the motor skills of potential pilots received an extraordinary amount of attention in the years surrounding and during WW II (Melton, 1947). A variety of electromechanical devices were used to simulate part of the pilot's job and then monitor performance on the task. The general scheme for selecting pilots involved collecting scored behavior samples prior to training, monitoring training outcomes and determining what probabilities of success were associated with the pretest scores. Unfortunately, from the validation standpoint, those doing poorly on the pretests were typically eliminated from further training, thereby preventing the determination of the extent of their predictive validity. In any event, the concept of judgment evaluation was not specifically addressed in that context.

Also during WW II, the Office of Strategic Services (OSS) needed to select individuals for specialized training and assignments. A personality factor considered to be important was called "Effective Intelligence," the "ability to select strategic goals and the most efficient means of attaining them; quick practical thought - resourcefulness, originality, good judgment - in dealing with things, people, or ideas" (The OSS Assessment Staff, 1947, p. 30). In addition to the gamut of questionnaires and projective tests, a systematic attempt was made to use practical tests in which candidates were placed in situations contrived to resemble ones they might face in field assignments.

Unfortunately, there were no criterion data with which to compare these procedures for validation. Nevertheless, the procedures were judged to be relatively successful, and their potential in the evaluation of

pilot judgment is evident. Situational tests are currently used to evaluate pilot skills during flight and, in view of the relative success with simulators used for this purpose by the airlines (e.g., Gibson, 1969), an investigation into systematic applications of situational tests for pilot judgment would appear to be a good investment.

Evaluating Pilot Judgment

Use of behavioral objectives as a major step in the evaluation of pilot judgmental training is relatively easy to implement. Flight instructors currently perform this function every day. The only change involves the increased specificity of the critical events. To insure the success of this type of evaluation it is necessary to provide training in observational techniques for flight instructors or anyone else involved in the formal assessment of pilot judgment. It is desirable in evaluating any human performance to avoid the pitfalls of both reactive and nonreactive measurements (Webb, Campbell, Schwartz, and Sechrest, 1966). Reactive measures occur whenever the subject is directly involved and is reacting to the measurement process itself. Nonreactive measures do not change that which is being measured. These measurements are generally passive and unobtrusive. The problem with nonreactive measures arises from sampling biases of the behavior. If the instructor is to remain the major instrument of measurement, it is imperative that his or her observations be as accurate as possible. The critical variable is instructor's ability to evaluate judgment, and it would appear that additional training would be needed to assure its adequacy.

Pretest Evaluation. Pretesting of the judgmental ability of individual candidates for private pilot certification is a potentially useful adjunct to the entire instructional and evaluation process. Pretesting used as a

baseline against which to measure the effects of instruction forms an essential component of the classic evaluation of program effectiveness. Ideally pretests would provide an opportunity for the individual to respond to the types of materials that would be used during training. Modification of these responses as a function of training would be taken as evidence of training effectiveness (assuming a similar group not exposed to the specific training manifested no such change). Because most beginning flight students lack the rudimentary knowledge to make appropriate judgments in all but the most obvious aviation situations, pretesting specifically for pilot judgment will likely be futile.

The role of the decision maker in a decision or judgmental process may be conceptualized as one step in the process. Figure 7 indicates that the decision maker acts on the situation with which he is confronted. Implicit within this model is the feasibility of evaluating decisions as both means and outcomes. More importantly, the model proposes that environmental characteristics such as the occasion for a decision and the organization of the information to be processed interact with the personal

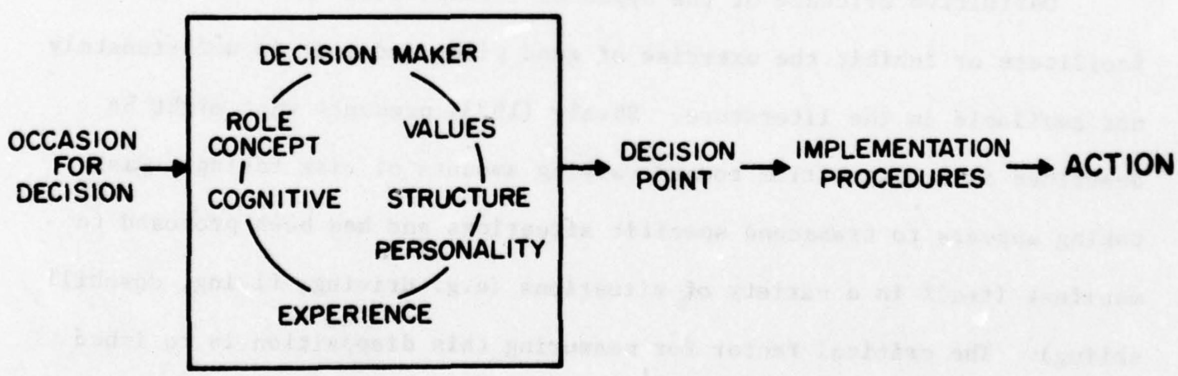


Figure 7. A simplified linear model of the decision process indicating that environmental characteristics (occasion for decision) and the organization of the information processing mechanisms interact with personal characteristics of the decision maker (adapted from Raser, 1969).

characteristics of the decision maker to produce judgments (Raser, 1969). While training procedures are testable in outcomes and, less frequently, means, pretesting can be accomplished by measuring such personal characteristics as the role concept, values, cognitive structure, personality, and experience. Experimental evidence is available to indicate that personal characteristics do bias individuals in the type of judgments they will prefer.

For example, Coombs and Pruitt (1960) demonstrated preferences for varying amounts of risk in gambling situations, where some subjects consistently preferred "long shots" and others played it safe. In a laboratory setting, Brody (1963) has shown personality variables of achievement motivation and anxiety to affect risk taking behavior. Taylor and Dunnette (1974) found dogmatism and intelligence to relate to risk taking during decisional behaviors. Additionally, the work during WW II on printed classification tests (Guilford and Lacey, 1949) indicates that there are measurable characteristics which the individual has acquired prior to pilot training which can affect the outcome of training.

Definitive evidence of the types of intrapersonal factors which can facilitate or inhibit the exercise of good pilot judgment is unfortunately not available in the literature. Shealy (1973) presents what might be described as a disposition toward varying amounts of risk taking. Risk taking appears to transcend specific situations and has been proposed to manifest itself in a variety of situations (e.g. driving, flying, downhill skiing). The critical factor for measuring this disposition is to imbed the test items within a specific situational context.

Most individuals who begin pilot training have had experience in automobile operation. Assuming Shealy is correct, much of the evidence applicable to automobile operation will be useful for comparison to aircraft

operation. Since a much larger body of literature exists on human factors in highway traffic safety, a review of this literature will serve as a model for directions for future work, pitfalls to avoid, and unfortunately, the complexity of the problem.

Skill and judgment form the heart of the driver's (or pilot's) task. Lack of skill and poor judgment undoubtedly are factors in the majority of accidents (Rockwell, 1972). As causal factors the two concepts cannot be easily separated since increases in one can often offset shortcomings in the other. Accepted analyses of the driving task have shown that the task is more than simple control manipulation and obstacle avoidance and the "more" is presumably judgment. To the extent that factors influencing judgment affect accident potential, the review of accident proneness literature may be important.

Suchman (1960-61) suggests we view accidents as the end product of a sequence of events which results in some unanticipated consequences that are judged as undesirable, placing the accent on the consequences. Therefore, medically speaking, accidents are considered as injury producing behavior. Nevertheless, the possibility exists that the resulting injury is a consequence of the unplanned event and does not itself constitute the accident -- it follows afterward.

An accident may be thought of as progressing through stages, possibly three broad segments. We have the actual event, preceded by the unsafe behavior in face of some existing hazard or danger, in turn preceded by some predisposing characteristic of the individual involved. Indeed it would be possible to define accidents merely in terms of the unsafe practices that might be considered to be instances of poor judgment. Unfortunately

for investigative purposes these seldom result in observable events that can be analyzed. Potential accidents that didn't happen would be the broadest possible conception of unsafe behaviors. However, the majority of literature focuses on actual injury producing events.

A popularly accepted factor in accident causation has been the concept of accident proneness. Accidents in a wide variety of situations have been attributed to accident proneness since Greenwood (Greenwood and Yule, 1920) reported his investigations of munitions factory accidents during WW I. The basic tenet of this concept is "...that even when exposed to the same conditions some people are inherently more likely to have accidents than others..." (Shaw & Sichel, 1971, p. 14). This concept is certainly well accepted by the majority of law officials and insurance companies. Individuals who manifest unusual numbers of accidents usually run into trouble from either or both (e.g. license revocations and insurance cancellations). However, scientific opinion is not as unanimous in its endorsement of accident proneness.

The major proponents of the modern interpretation of accident proneness have been Shaw and Sichel (1971). They maintain that accidents are more than random events, but accident prone individuals do not form a unique class of their own. This implies that there will not only be unequal involvement in accidents, but also unequal accident potential. According to Shaw and Sichel, it is the individuals who for lack of skill, physical defects, personality defects, or incorrect attitudes have a dangerous amount of accident potential. Furthermore, under conditions of strict supervision this potential would manifest itself only occasionally in poor vehicle manipulation or the occasional accident or near accident, but would show itself far more clearly if the supervision were less strict.

An important implication of this position is that there is probably not a specific minority of individuals, i.e. five percent or so, who have the majority of accidents. There is an element of accident proneness that is present in nearly everyone but is suppressed in some while developing into an actual disposition in others. The degree to which accident proneness is displayed by a particular person in a particular situation is primarily determined by a combination of his own structural personality characteristics and the specific situation. Additionally, what a person does or neglects to do in a given situation is determined initially by his accident potential.

Haddon, Suchman, and Klein (1964) present what may be termed the "folklore view", that somehow accidents occur (e.g. as acts of God) and they mysteriously defy any kind of systematic study other than mere tabulation. Haddon, et al. also note that a great many people have their own pet theory of how to solve the accident problem. Both of these views have pervaded the accident research literature and are at least partially responsible for the recent lack of support for the concept of accident proneness. The most telling criticisms come from those statistical investigations of accidents which fail to support the concept. McFarland (1962) states, "It has not been convincingly demonstrated that an appreciable number of people tend to have more accidents than others under conditions of equal exposure." It has been this apparent inability to find a small percentage of the drivers who have been having the majority of the accidents and the apparent failure to find the strong correlations between individual personal attributes and accident causation that has disillusioned a large number of investigators. For those who have not abandoned the concept of accident proneness entirely there is guarded acceptance of a limited form of the concept.

One form of this acceptance involves an acknowledgement that nothing

in life is simple and indeed the hope for a single causative agent in accident research exceeds reality. McFarland and his colleagues (e.g. McFarland, Moore and Warren, 1955) present considerable evidence as to the pertinence of various attributes of individuals. Recognition of the possibility of a multifaceted form of accident proneness is essential for scientific investigation. Another possibility is that accident proneness endures for varying lengths of time and some individuals are definitely accident-prone but for only short durations (McGuire, 1976).

One approach to psychological fitness for driving relates to an exploration of the person's lifestyle. The basic belief is that personal maladjustment will be reflected in driving behavior (and presumably any other vehicle manipulation). In a preliminary study, Tillman (in McFarland, 1958), found high frequencies of various adverse background events (e.g., divorced parents, truancy, poor employment record) in accident repeaters. Following this line of reasoning, a 15 year history of 2 accident free groups and one accident repeat group was made. The results indicate that accident repeaters have had a great deal more contact with the listed agencies (adult court, juvenile court, social services, public health department, etc.). The accident repeater was known to one or more agencies in 66 percent of the cases as compared to only 9 percent for the accident-free group.

This line of research is not without critics. Hacker and Suchman (1963) point out that these studies, in general, have methodological deficiencies. Many resemble fishing expeditions lacking clearly formulated hypotheses to explain the relationship between indepedent and dependent variables. Others lack control groups and merely indicate that accident repeaters have certain common factors. When control groups were used, inadequate matching on crucial variables occurred (e.g. experience and exposure). The fault of

many of these studies also relates to the overinterpretation of differences between extreme groups. Little, if any, attention is paid to the nature of the population distribution and size of the overlooked middle, when estimates of the magnitude of the relationship are made.

Another more controversial approach uses the assumed relationship between attitudes and accident and/or violation involvement. Some researchers have maintained that there are no homogenous clusters of attitudinal variables that can be used to discriminate the good from the bad driver. Others, who have supported the importance of attitudinal variables, will admit that many of the observed relationships are weak, and that the outcomes of most correlational studies have been disappointing. Yet, there are researchers who have found significant correlations in the accident-attitude relationship.

If attitudes and other intrapersonal attributes are accepted as causative agents in accidents (and accidents are primarily due to faulty judgments), then testing for these attributes should be an essential part of the licensing procedure. Unfortunately, at this time most paper and pencil tests of personality, etc. are capable at best of exposing only those with the grossest deficiency. Blanket standards (which would screen unfit individuals) would be difficult to propose with such unwieldy instruments (Kimball, 1970). The most beneficial use of such tests would be seen in screening for the chronically negligent drivers who continuously accrue violations and accidents for which they are responsible. Goldstein (1964) has noted the extremes on variables such as personality, emotion, attitudes and socioeconomic status may be valid criteria for prediction of driver behavior, but reliability of such tests on these variables remains to be demonstrated.

The study of attitudes per se has not been ignored. Goldstein and Mosel (1958) when studying the identification of the dimensions which underlie driver

attitudes, used clusters of attitudes from which 186 items were drawn. From these items, five factors were identified:

Attitude toward competitive speed
Attitude toward other users of the roadway
Attitude toward the "cops"
Attitude toward the vehicle
A general attitude of care or concern for safety

Although the authors had not validated these for their ability to predict accidents, supporting evidence for these is available from other sources. Rainer, Conger, Gaskill, Glad, Sawrey, Turrel, Walsmith, and Keller (1959) found that persons who are accident repeaters do tend to be more nonconforming in their opinion and attitudes as well as their social conduct and at the same time were more tense, less able to handle tension, and therefore more prone to act out emotional conflicts and involvements than non-accident-involved persons.

A report from Kainuma (cited in Lucas, 1970, p. 138) in Japan deals with impulsiveness and accidents. This report reveals that accident repeaters were more irritable and more prone to act impulsively than non-accident drivers. In specific psychomotor test combinations, Kainuma found no ability of simple and choice reaction time to discriminate, but a third test involving a more complex combination of both perceptual discrimination and psychomotor skill showed accident-free drivers to be superior. The report of personality factors indicates accident repeaters had a variety of undesirable attributes associated with them, including being careless in personal matters, hedonistic, displeased with family life, etc. However, Kainuma never claims to be able to predict future accident repeaters or to discriminate accident prone drivers from good ones.

Given the preceding evidence that there are at least some characteristics which separate at least some of the accident repeaters from good drivers, it

is not surprising that a number of tests have been designed which attempt to discriminate between good and bad drivers. The "Driving Survey" (Case and Stewart, 1957), the "Driver Attitude Survey" (Schuster and Guilford, 1962) and the "McGuire Safe Driver Scale" (McGuire, 1961) have all had modest success in predicting accident involvement. None of these are sufficient in themselves to form the basis of a decision to grant or not grant licensing to an individual driver. Ultimately, of course, development of predictive devices based on drivers' attitudes should be the goal of researchers in this area. Unfortunately, few firm conclusions can be drawn from the best of the available research in this area, and a considerable body of less adequate research exists to cloud the issue.

WW II saw an increase in research relevant to pilot errors or lack of good judgment as a cause of accidents. Kalez and Hovde (1945) reviewed the records of pilots and report what they call a unique psychological group of pilots who willfully failed to use checklists. They refer to this temporary psychological compulsion as an error in judgment. The pilots who were involved in the resulting accidents presented a long history of nonconformity as evidenced from their flight records. Occasionally they were below average aviators in every measurable respect. The type of errors in which they were interested were classified as "pilot errors". In their opinion, the "accident prone" aviators belonged to this group. The accidents resulted from an initial error of judgment which was then coupled with poor technique. They reported 18.4 percent of the pilots caused 58.2 percent of the pilot inaptitude accidents. They found five factors valid in the production of repeated accidents. In 23 percent of the pilots a below average flight training record was found. The fatal crash rate was 3.4 times that of pilots with average and above average flight training records. There also appeared

to be below average adaptability to flight, but it was not clear whether it was present prior to training or occurred after training began. An excessive grounding history appeared in 31 percent of pilots in question. In 9 percent of the pilots, a history of "doping off" in the air was present, where lapses of memory concentration and attention initiated a dangerous aviation incident (not an accident). As education standards applied in military pilot selection were lowered, those with less education tended to be overrepresented in the accident group.

The nature of "pilot error" was further investigated by Kunkle (1946). The term refers to all the defects which a pilot may exhibit in all aspects of aviation, although it is limited in that it primarily refers to accident causing behavior. He divided the topic into primary or "non-emergency" error including errors in judgment and secondary or "emergency" error occurring in a setting of tension and confusion associated with a crisis in some phase of flight. Of course, the secondary error may result from an immediately preceeding pilot error. However, true emergencies were responsible for only a very small minority of pilot error accidents. Kunkle reports the general failing of the then current (WW II) selection tests for prediction of pilot error accidents and concomitantly an inability to assess an individual's judgment. He further implies that the problems involved in the operation of an airplane allow full scope for any of the various manifestations of accident proneness. Reviewing the accident background of pilots involved in accidents revealed a previous accident pattern in which auto accidents were indeed conspicuously frequent. He concludes that there is a direct, but by no means, rigid, relationship between past performance on the ground and safety record in the air. Kunkle feels this relationship would be greater if the sample were less restricted (only rated pilots were included, therefore considerable selection had already occurred through processes of attrition, suspension, or death).

In the work on printed classification tests for aircrew members, Guilford and Lacey (1949) report that of all the traits necessary for pilots, judgment stands out as being the most persistent and universal. However, the frequent mention of judgment for the pilot presented a persistent challenge to break it down to manageable components and devise tests for it. In 1940, Kelly (Guilford and Lacey, 1949, p. 124) had flight instructors rate student pilots and found the following items from a 14 item checklist to be related to a judgment factor:

> How good is his judgment with regard to taking flying risks (weather, stunting, etc.)?
>
> Does he show respect for a ship and its motor?
>
> How well is he satisfied with his flying ability?
>
> Is he inclined to show off while flying a plane?
>
> How carefully does he check his plane and engine before taking off?

The work indicating that practical judgment was a part of safe flying led to further investigations of both judgment tests and instructor reports of practical judgment.

Guilford and Lacey (1949) concluded that the judgment factor was best defined by a work-planning type of item. Further, items calling for relatively complicated estimates involving time as well as distance and size were significantly loaded with the judgment factor, while the simpler items were not. The inference which may be drawn is that the judgment factor is a thinking, rather than a perceptual or memory, ability. Guilford concludes that judgment was very highly regarded as a factor and received considerable attention (with varying degrees of success). A promising avenue which opened at the end of the war was the hypothesis that thought fluency might be an independent contributor to practical judgment. The facile recall of pertinent experiences

for possible use in everyday predicaments would give an individual more potential solutions from which to choose. Apparently the construction of such tests using answer sheets presented some problems, but it was suggested that this hypothesis should have been followed up if possible. In addition, it might be profitable to collect, from training instances of good and poor judgment, the types of information that would have been pertinent to a successful solution of the problem.

The measurement of the adequacy of judgment is a difficult task. Rockwell (1972) attributes this difficulty to the observer's inability to adequately evaluate the judgment without knowing the alternative courses of action open to the driver and his motivation behind a given decision. According to Rockwell, judgments in automobile driving should be divided into three categories:

1. Emergency decisions -- one time situations requiring unique often rapid decision-making and response. Simulation of these events lacks the realism of real-world threats and repeated exposures destroy the uniqueness. Field studies are too dangerous and again limit coverage to a single exposure.

2. Decisions whether or not to engage in driving -- the situation where the driver is tired, drinking, aware of vehicle defects, etc. These are also unique decisions and not easily researchable although perhaps the key to good driver judgment.

3. Operational decisions, including routine decision as to speed selection, passing and merging decisions. These decisions are more amenable to observation and research and are important because they dictate the level of subsequent skill required of the driver to execute the decision and maintain a given performance.

A definite similarity exists between the judgments listed above and those enumerated earlier as behavioral objectives for flight training. Given this similarity, a question of interest is: to what extent are the operators of these vehicles capable of meeting these decision demands? Cumming (1964) has suggested that there is an upper limit to the rate at which the operator can make decisions. Obviously the decisions that an operator must make are many and varied, ranging from minor automatic decisions to highly complex decisions relating to potential accident situations. Not only do decisions vary widely, they also differ by type. Some decisions are dichotomous (e.g., passing, merging, or stopping) while others are continuous, (e.g. vehicle control inputs). It is not apparent from the literature how individual differences in the role of decision-making may account for accidents, but the research in this area is also needed.

In a recent review of accidents and risk, Taylor (1976) discusses what he calls the intentional aspects of risk. He notes that it is not a logical impossibility to take a risk and not intend to suffer any untoward consequences. Certainly the majority of risks are undertaken with the belief that there is a high probability of success. It is important to remember that the subjective impression of danger, regardless of how miniscule, or whether actual, is a necessary condition to define risk. Tracing the intended risk taking behavior to an accident dictates that at some point the events leave the control of the involved individual. Near accidents might now be viewed as instances where control was regained prior to the accident. Taylor proposes that maintenance of control is the essence of <u>safe</u> behavior. Control is defined as the power to direct the machine in accordance with intention. This removes one ambiguity but leaves the unobservable intention.

The driver's opportunity to perceive loss of control and this perception is subjective risk (Taylor 1976).

The critical aspects of a situation are the perceived risk and the amount of risk an individual is willing to accept. Increases in safety of a system may be achieved by many routes. One possibility is to increase the perceived risk involved with a given course of action. If the acceptable risk level remains constant, increases in perceived risk may cause the decision maker to act in a more cautious manner. Naatanen and Summala (1974) have shown that making the environment safer in both appearances and actuality has eliminated the expected safety gains. For example, better road surfaces allowed increases in both speed and frequency of passing up to the new level of subjective risk thereby attenuating the effects of subjective risk. It is assumed that subjective risk increases are reflected by increased vigilance. The benefits of this evidence to the previously presented model of Janis and Mann (1977) are apparent.

Direct determination of risk taking tendencies would be particularly useful if the demonstrated relationships apply to aircraft operation as well. Gumpper and Smith (1968) attempted to find a relationship between two paper-pencil measures of risk-taking tendencies and accident records in truck drivers. Neither the Life Experience Inventory (designed by Torrance and Ziller, 1957) nor Wallach & Kogan's (1959) Dilemmas of Choice questionnaire were able to discriminate between high and low accident drivers in Gumpper and Smith's sample. Although these results are not encouraging, the possibility exists that risk taking tendencies may be found through other measures. It is also possible that these measures would have greater success when applied to different criterion groups. The difference between high (89 accidents per year) and low (.12) groups was very small. This topic appears to be promising and warrants research beyond this one study.

Although the evidence is not definitive, it appears that the investigation of intrapersonal factors as correlates of poor judgment would be a project with considerable merit. Several very promising approaches to the study of intrapersonal factors have not been fully exploited. Perhaps the most promising of these is the situational approach of Shealy (1973). Use of situational tests encompassing the existing risk and decision-making research should have the greatest potential payoff. To the extent that an individual's attitudes will affect the quality of that person's judgments, then personal attitudes would be a useful index. Although not really test data, the biographical history of individuals has frequently been mentioned as a correlate of accident occurrence. The individual with the adverse background may also be shown to hold attitudes consistent with that background and this combination may be sufficient to contraindicate positions of responsibility. Only well-conducted research specific to aviation will resolve the conflicting results reported from the diverse sources.

The question of how to use pretest data must be considered discretely in our free society. During and following WW II, selection tests were used by the military as a screening process to lower the attrition rate from primary flight schools. One major problem when attempting to provide a better fit between man and machine, and hence lowered accident probability, applies to the task of selection. According to Klein (1976), it has been assumed that: (1) a large pool exists from which operators may be selected who embody the proper characteristics; (2) these operators could be forced to behave appropriately, by threats or economic incentives; and (3) performance of the man-machine system could be continuously monitored. Klein admits these assumptions probably hold for the professional operator, e.g. commercial airline pilots or interstate bus drivers. But, Klein quickly indicates

that such restrictions may not be applicable to the more casual operator. In the amateur situation, less rigorous selection procedures are in force, economic sanctions are not easily enforced and continuous monitoring is essentially impossible. Although the potential for a rigorous selection approach exists in civil aviation, a more acceptable purpose of pretest data would be to improve pilot training procedures. Individuals might not be screened out, but identified as needing greater attention in judgmental training.

These issues are what Cronback and Gleser (1965) termed institutional decisions. When an individual considers one person alone (e.g. approving him for private pilot certification), in applying the value system of society, he is making an institutional decision. The decision in this case is made to minimize the loss from an erroneous decision. It will not always be the case that what is best for the individual will be best for society. Any testing authority is forced to consider the ramifications of decisions based on one or a series of tests. A balance between individual freedom and societal protection is needed.

Although the question of how far into an individual's background regulatory agencies should probe is beyond the scope of this report; this question is important if one accepts such statements as "Man drives as he lives." (Tillman and Hobbs, 1949). It may be said that the entire adjustment complex of the individual reflects the accident record of this person as a vehicle manipulator. The person who makes continual mistakes in facing the demands of the social and personal milieu is also likely to err in the performance of his obligations in vehicle manipulation. This concept has been shown to be basically true, as reflected in the financial, social and employment records of a sample of violators and accident-involved persons (Lucas, 1970).

For truly effective implementation of pilot judgment selection and instruction, a more intensive background investigation would appear to be an essential component to the entire process.

Training Evaluation. Assuming the use of behavioral objectives is implemented successfully, monitoring and evaluating training become matters of noting progress toward the stated objectives. The precise nature of the evaluation will vary with the type of instruction in use. If a section of conventional ground schools were devoted to judgment training, the ultimate determination of the effectiveness of this training would likely be the conventional end of course examination. The integrated use of innovative techniques such as interactive movies would allow a direct check on performance through recorded responses to the movie. The use of role-playing within the conventional classroom presents a unique opportunity for evaluation. It would allow the classmates to evaluate the judgments exercised by the student playing pilot, and it would allow the instructor to note these additional evaluations. This resembles a simulation with test items imbedded in the situation (Shealy, 1973) and should be a very effective technique for evaluating judgmental instruction. A further benefit from role-playing not available in written exams is the ability to evaluate the process for arriving at the judgment, and not just the end product.

Although the airplane is not necessarily the ideal environment for instruction and subsequent evaluation, it is obviously an integral part of pilot training programs. Instructors often have a great deal to do without the additional demands of systematic evaluations. A form of shorthand would alleviate much of the burden. The critical incident (CI) technique is a set of procedures for collecting direct observations of human behavior in such a way as to facilitate their potential usefulness of solving practical problems and developing broad psychological principles (Flanagan,

1954). The technique outlines procedures for collecting observed incidents having special significance which meet systematically defined criteria. An incident is defined as any observable human activity that is sufficiently complete within itself to permit inferences and predictions about the person performing the act. Critical refers to the purpose of intent of the act which must appear clear to the observer and whose consequences are sufficiently definite to leave little doubt concerning the effect.

The essence of the technique is that only simple types of evaluations are required of the observer. Reports are only taken from qualified observers and all observations are evaluated by the observers in terms of an agreed upon statement of the purpose of the activity. Currently, this CI technique is implemented in a more or less systematic way by most flight instruction and evaluation organizations. When the CI technique is used to measure typical performance (criteria), a simple checklist on which to record tallies for instances of the observed behavior is a basic application. For measures of proficiency (e.g. a standardized sample such as a flight check) the CI technique provides uniformly standardized opportunities to perform the critical aspects of the job.

While the CI technique is currently used in its essential form, the more refined aspects of the technique are not always adhered to. For selection and classification (and certification) purposes, a more thorough study of the job under investigation is necessary to avoid the so-called "shotgun approach". Implicit within the CI framework is a carefully considered analysis of the essential features of the pilot's job and knowledge of likely problem areas. Although it may seem that many current programs use this technique, on closer examination, it may be revealed that only cursory evaluation is given to some aspects of flight while others are overrepresented.

Also, periodic reanalysis of currently used procedures is essential to determine if any changes have occurred in the critical features of the pilot's job.

The major benefit of a critical incident technique is the determination of unsafe procedures prior to the occurrence of an accident. This benefit is especially important in avoiding the pitfalls inherent when attempting to assign blame for accidents after the fact. Errors in pilot judgment have been shown to be a major factor in aviation accident causation. However, our knowledge of events immediately preceding the accident is exceedingly limited. For example, it is insufficient for purposes of accident prevention to state merely that the pilot in command failed to exercise good judgment and in the future urge all pilots to exercise good judgment. Recording of near accidents will allow both student and instructor to examine these events in detail and learn from the experience.

Tarrants (1965) has shown that the CI technique dependably reveals errors and/or unsafe conditions which lead to accidents. More importantly, it can identify causal factors associated with both injurious and non-injurious accidents. The CI technique can also reveal a greater amount of information about accidents than more widely used current methods of accident study and can provide a more sensitive measure of total accident performance. Lastly, causes of non-injurious accidents identified by the CI technique provide identifiable sources of potentially injurious accidents. This means that failures in judgment which have no untoward consequences in one situation could, with very slightly altered circumstances, have resulted in an accident of catastrophic proportion in another.

Rockwell, Bhise, and Clevinger (1970) performed a study in which the CI technique was used to assess the safety of flight operations in a tactical

fighter squadron. The study showed that the application of the CI technique was a useful method of finding pilot errors. Additionally, the fact that the pilots rated themselves for safety after every mission made them aware of what is unsafe everytime they completed a checklist. This increased awareness was reflected by the greater degree of safety evident in the later weeks of the study. The authors conclude that the greatest advantage of the technique is the availability of continuous assessment of pilot performance and consequently, helps in evaluating safety programs. The feasibility of application to a pilot judgment instruction program is self-evident.

The systematic evaluation of judgment during the training program has a potentially valuable side benefit. Awareness of the behavioral criteria of good judgment should lead to increased attention to monitoring one's own judgmental behavior. Although experimental evidence is lacking, an increase in good judgment following training should be expected. The transfer of training to the operational environment is an important aspect of the program's effectiveness.

Transfer Evaluation. Evaluation of transfer in a judgmental training program is not as easily implemented as the pretest or training evaluation. Following training, much of the direct control over the individuals to be evaluated has been lost. Criterion data to determine effectiveness depend to a large extent on those statistics indicative of poor judgment. Over a long term, NTSB statistics would provide evidence of success of transfer. An adaptation of the transfer effectiveness ratio (Roscoe, 1971; 1972) would provide a basis for comparison of accident rates for those specially trained in judgment against control pilots who have experienced conventional flight training. With effective transfer, those trained in judgment would have a lower judgment-related accident rate than pilots without specific judgment

training, and this difference could be expressed as a modified transfer effectiveness ratio.

If lower accident rates (reflecting the value of judgment training) were found in the trained group, it would be instructive to check members of this group periodically, using other measures, to determine their current effectiveness in rendering good judgments. This would allow investigation of the factors from training that are retained. The increased awareness of rendering judgments, specific procedures from training, or ability to attend to relevant variables while ignoring the irrelevant are some of the factors that might be investigated in such an evaluation. Input from these transfer evaluations could be returned to the program to help modify the program for greater effectiveness.

The potential benefit of computer-assisted instruction to pilot judgmental evaluation was mentioned earlier. Using CAI a number of judgmental areas could be evaluated and refreshed if necessary. For example, in the context of an examination, the student could be presented with a flight situation requiring judgment. He could then be asked to respond by listing the relevant information required, the alternatives available, the factors affecting each alternative, and the probability of success for each alternative, and the probability of success for each alternative for his own flight experience level. An examination of student responses in comparison with previously stored alternatives in several situations could provide the basis for an evaluation of the pilot's judgment. Data from these evaluations could be stored temporarily or permanently for use in improving the test programs through comparison with actual judgmental performance in flight.

What Evaluation Can Mean

Evaluation in a pilot judgmental training program need not only be a method of determining student progress. An analogy may be drawn to standardized achievement tests. While standardized tests provide feedback to each individual student, they also tell the teachers and school system how they compare to other individual or reference groups. A system of judgment evaluation could indicate to flight instructors areas in which their instructive skills are sufficient and where they are not. It could perform the same function for an entire program.

CONCLUSIONS

A. Pilot judgment can be and is defined as follows:

1) The ability to search for and establish the relevance of all available information regarding a situation, to specify alternative courses of action, and to determine expected outcomes from each alternative.

2) The motivation to choose and authoritatively execute a [redacted]e course of action within the time frame permitted by the situation.

Where: (1) "Suitable" is an alternative consistent with societal norms; (2) "Action" includes no action, some action, or action to seek more information.

B. Pilot judgment is not now, in any structured manner, being evaluated in civil flight training, but evidence described herein supports the conclusion that such evaluation is feasible.

C. Training pilots to demonstrate good judgment is feasible and a method for doing so is broadly outlined herein.

REFERENCES

Abeles, H. F. A facet-factorial approach to the construction of rating scales to measure complex behaviors. Journal of Educational Measurement, 1973, 10, 145-151.

Adams, J. A. Learning and memory: An introduction. Homewood, IL: The Dorsey Press, 1976.

Alpert, D. and Bitzer, D. L. Advances in computer-based education. Science, 1970, 167, 1582-1590.

Anastasi, A. Psychological testing. 3rd ed. New York: McMillan, 1968.

Anderson, N. H. Comment on "analysis-of-variance model for assignment of configural cue utilization in clinical judgment." Psychological Bulletin, 1969, 72, 63-65.

Asch, S. E. Effects of group pressure upon the modification and distortion of judgments. In H. Gretzkow (Ed.) Groups, leadership and men: Research in human relations. Pittsburg: Carnegie Press, 1951, 177-190.

Atkinson, J. W. Motivational determinants of risk-taking behavior. Psychological Review, 1957, 64, 359-372.

Aviation Consumer interview with Webster Todd March 1, 1977, Vol VII.

Bellows, R. Psychology of personnel in business and industry, (3rd ed.). Englewood Cliffs, N.J.: Prentice Hall, 1961.

Berrien, F. K. A general systems approach to organizations. In M. D. Dunnette (Ed.) Handbook of Industrial and Organizational Psychology, Chicago: Rand-McNally, 1976, p. 43.

Birdsall, T. G. The theory of signal detectability. In H. Quastler (Ed.) Information theory in psychology: Problems and methods, Glencoe, IL: Free Press, 1955, pp 391-402.

Borg, W. R. Educational research: An introduction. New York: David McKay Co., 1963.

Brody, N. N achievement, test anxiety, and subjective probability of success in risk taking behavior. Journal of Abnormal and Social Psychology, 1963, 66, 413-418.

Brunswick, E. Perception and the representative design of psychological experiments. Berkeley, CA.: University of California Press, 1956.

Caro, P. W. Aircraft simulators in pilot training. Human Factors, 1973, 15, 503-510.

Case, H. W. and Stewart, R. G. The development of a driver attitude scale. Washington, D.C.: Highway Research Board Bulletin, 172, 1957.

Chernoff, H. and Moses, L. E. Elementary Decision Theory. New York: Wiley, 1959.

Cooley, W. W. and Glaser, R. The computer and individualized instruction. Science, 1969, 166, 574-582.

Coombs. C. H. and Pruitt, D. G. Components of risk in decision making: Probability and variance preferences. Journal of Experimental Psychology, 1960, 60, 265-277.

Cronbach, L. J. and Gleser, G. C. Psychological tests and personnel decisions. Urbana, IL: University of Illinois Press, 1965.

Cumming, R. W. The analysis of skills in traffic. Australia Road Research. Victoria, Australia, 1964.

Damos, D. L. Development and transfer of time-sharing skills. Urbana, IL.: University of Illinois at Urbana-Champaign, Unpublished doctoral dissertation, 1977; also Savoy, IL: University of Illinois at Urbana-Champaign, Institute of Aviation, Aviation Research Laboratory, Technical Report ARL-77-11/AFOSR-77-10, July, 1977.

Deutsch, M. and Gerarard, H. B. A study of normative and informational social influences upon individual judgment. Journal of Abnormal and Social Psychology, 1955, 51, 629-636.

Edwards, W. and Tversky, A. (Eds.) Decision making. Baltimore: Penguin Books, Inc., 1967.

Federal Aviation Administration. Aviation Instructor's handbook. Washington, D.C.: U.S. Government Printing Office, No. AC 61-14, 1977.

Fine, P. M. and Hartman, B. O. Psychiatric strengths and weaknesses of typical Air Force pilots. Brooks AFB, TX.: USAF School of Aerospace Medicine, 1968.

Finnegan, J. P. Evaluation of the transfer and cost effectiveness of a complex computer-assisted flight procedures trainer. Urbana, IL.: University of Illinois at Urbana-Champaign, Unpublished master's thesis, 1977; also Savoy, IL: University of Illinois at Urbana-Champaign, Institute of Aviation, Aviation Research Laboratory, Technical Report ARL-77-7/AFOSR-77-6, June 1977.

Flanagan, J. C. The critical incident technique. Psychological Bulletin, 1954, 51, 327-358.

Flexman, R. E., Roscoe, S. N., Williams, A. C., Jr., and Williges, B. H. Studies in pilot training: The anatomy of transfer. Aviation Research Monographs, 1972, 2 (1).

Forbes, T. W. The normal automobile driver as a traffic problem. Journal of General Psychology, 1938, 20, 471-474.

Frank, L. K. Projective methods. Springfield, IL.: Thomas, 1948.

Gagne, R. M. Military training and principles of learning. American Psychologist, 1962, 17, 83-91.

Galanter, E. and Holman, G. L. Some variations of the insensitivity function and their implications for the utility of money. Journal of Experimental Psychology. 1967, 73, 333-339.

Gibson, J. H. Optimized flight crew training: A step toward safer operations. Fort Worth, TX.: American Airlines, Flight Training Academy, 1969.

Gilbert, T. F. On the relevance of laboratory investigation of learning to self-instructional programming. In A. A. Lunsdaine and R. Glaser (Eds.), Teaching machines and programmed instruction. Washington, D.C.: National Education Association, 1960.

Glaser, R. (Ed.). Training research and education. Pittsburg: University of Pittsburg Press, 1962.

Goldberg, L. R. Simple models of simple processes? Some research on clinical judgments. American Psychologist, 1968, 23, 483-496.

Goldstein, I. L. Training: Program development and evaluation. Belmont, CA.: Wadsworth, 1974.

Goldstein, L. G. Human variables in traffic accidents: A digest of research. Traffic Safety Research Reviews, 1964, 8, 26-31.

Goldstein, L. G. and Mosel, J. N. A factor study of driver's attitudes with further study on driver aggression. Washington, D.C.: Highway Research Board Bulletin, 172, 1958.

Greenwood, M. and Yule, G. U. An enquiry into the nature of frequency distributions representative of multiple happenings, with particular reference to the occurrence of multiple attacks of disease or repeated accidents. Journal of the Royal Statistical Society, 1920, 83, 255-279.

Guilford, J. P. and Lacey, J. I. Printed classification tests. Washington D.C., U.S. Government Printing Office, 1947. (AAF Aviation Psychology Program Research Report No. 5)

Gumpper, D. C. and Smith, K. R. The prediction of individual accident liability with an inventory measuring risk-taking tendency. Traffic Safety Research Review, 1968, 12(2), 50-55.

Hacker, H. and Suchman, E. A. A sociological approach to accident research. Social Problems, 1963, 10, 383-389.

Haddon, W. Suchman, E. A. and Kline, D. (Eds.) Accident research: Methods and approaches. New York: Harper and Row, 1964.

Harshbarger, T. L. Pilot education: The search for a better way. The Flight Service Journal, 1967.

Hendrikson, G. and Schroeder, W. Transfer of training in learning to hit a submerged target. Journal of Educational Psychology, 1941, 32, 206-213.

Hoffman, P. J. The paramorphic representation of clinical judgment. Psychological Bulletin, 1960, 57, 116-131.

Hoffman, P. J., Slovic, P., and Rorer, L. G. An analysis of variance model for the assessment of configural cue utilization in clinical judgment. Psychological Bulletin, 1968, 69, 338-349.

Jacobs, R. S. and Roscoe, S. N. Simulator cockpit motion and the transfer of initial flight training. Dallas, TX.: Proceedings of the Nineteenth Annual Meeting of the Human Factors Society, October 1975.

Janis, I. and Mann, L. Coping with decisional conflict. American Scientist, 1976, 64 (6), 657-667.

Janis, I. and Mann, L. Decision making: A psychological analysis of conflict, choice, and commitment. New York: Free Press, 1977.

Johnson, C. D. and Davis, J. H. An equiprobability model for risk-taking. Organizational Behavior and Human Performance, 1972, 8, 159-175.

Kalez, M. M. and Hovde, R. C. Pilots with repeated "pilot-inaptitute" accidents. Journal of Aviation Medicine, 1945, 16, 370-375.

Kimball, K. A. Driver licensing and selection: Current shortcomings and research issues. In N.W. Heinstra, ed. Injury control in traffic safety. Springfield, IL.: Thomas, 1970.

Klein, D. Social aspects of exposure to highway crash. Human Factors, 1976, 18, 211-219.

Kogan, N. and Wallach, M. Risk taking: A study in cognition and personality. New York: Holt, Rinehart, and Winston, 1964.

Kunkle, E. C. The psychological background of "pilot error" in aircraft accidents. Journal of Aviation Medicine, 1946, 17, 533-567.

Lee, W. Decision theory and human behavior. New York: Wiley, 1971.

Levine, H. G. and McGuire, C. Role playing as an evaluative technique. Journal of Educational Measurement, 1968, 5, 1-8.

Levine, H. G. and McGuire, C. The validity and reliability of oral examinations in assessing cognitive skills in medicine. Journal of Educational Measurement, 1970, 7, 63-74.

Liebert, R. and Morris, L. Cognitive and emotional components of test anxiety: A destination and some initial data. Psychological Reports, 1967, 20, 975-978.

Lord, F. M. and Novick, M. R. Statistical theories of mental test scores. Reading, MA.: Addison Wesley, 1968.

Lucas, R. L. Attitudes, personal characteristics, and driver behavior. In N. W. Heimstra, ed. Injury control in traffic safety. Springfield, IL.: Thomas 1970.

Mager, R. F. Preparing instructional objectives. Palo Alto, CA.: Fearon, 1962.

McFarland, R. A. Human Factors in highway transport safety, Washington, D.C.: Highway Research Board Bulletin 60, 1958, 36-43.

McFarland, R. A. The epidemiology of motor vehicle accidents. Journal of the American Medical Association, 1962, 180, 289-300.

McFarland, R. A., Moore, R. C., and Warren, A. B. Human variables in motor vehicle accidents: A review of the literature. Harvard School of Public Health, 1955.

McGuire, F. L. The McGuire safe driver scale. Beverly Hills, CA.: Western Psychological Services, 1961.

McGuire, F. L. Personality factors in highway accidents. Human Factors, 1976, 18, 433-441.

McGuire, C. and Babbott, D. Simulation techniques in the measurement of problem solving skills. Journal of Educational Measurement, 1967, 4, 1-10.

McMurris, R. N. Validating the patterned interview. Personnel, 1947, 23, 263-272.

Melton, A. W. (Ed.) Apparatus tests. Washington, D.C.: U.S. Government Printing Office, 1947. (AAF Aviation Psychology Program Research Report No. 4)

Naatanen, R. and Summala, H. Physical and psychological aspects of crash barriers. Accident Analysis and Prevention, 1973, 5, 247-251.

The OSS Assessment Staff, Assessment of men: Selection of personnel for the Office of Strategic Services. New York: Rinehart, 1947.

Parducci, A. The relativism of absolute judgments. Scientific American, 1968, 219(6), 84-90.

Pelz, D. C. Driving "immunization" in alienated young men. Human Factors, 18, (5), 1976, 465-476.

Pirsig, R. M. Zen and the art of motorcycle maintenance. New York: Bantam, 1975.

Povenmire, H. and Roscoe, S. The incremental transfer effectiveness of a ground-based general aviation trainer. Human Factors, 1973, 15, 537-545.

Rabin, A. I., (Ed.) Projective techniques in personality assessment: A modern introduction. New York: Springer, 1968.

Rainey, R. V., Conger, J. J., Gaskill, H. S., Glad, D. D., Sawrey, W. L., Turrell, R. S., Walsmith, C. R., and Keller, L. An investigation of the role of psychological factors in motor vehicle accidents. Washington, D.C.: Highway Research Board Bulletin, 1959.

Raser, J. R. Simulation and society: An exploration of scientific gaiming. Boston: Allyn and Bacon, 1969.

Rockwell, T. Skills, judgment, and information acquisition in driving. In T. W. Forbes, ed. Human factors in highway traffic safety research, New York: Wiley, 1972.

Rockwell, T. H., Bhise, V. D. and Clevinger, T. R. Development and application of a non-accident measure of flying safety performance. Journal of Safety Research, 1970, 2, 240-250.

Roscoe, S. N. Galileo and the marketing manager. Aviation Research Monographs, Savoy, IL: Institute of Aviation, University of Illinois; 1971, 1, 1-9.

Roscoe, S. N. Incremental transfer effectiveness. Human Factors, 1971, 13, 561-567.

Roscoe, S. N. A little more on incremental transfer effectiveness, Human Factors, 1972, 14, 363-364.

Roscoe, S. N. Man as a precious resource: The enhancement of human effectiveness in flight operations. New York, N.Y.: AIAA Paper No. 74-1296, Nov. 1974.

Rotter, J. B. General expectations for internal versus external control of reinforcement. Psychological Monographs, 1966, 80(1).

Rubinsky, S. and Smith, N. Safety training by accident simulation. Journal of Applied Psychology, 1973, 57, 68-73.

Savage, L. J. The foundations of statistics. New York: Wiley, 1954.

Schuster, D. H. and Guilford, J. P. The driver attitude survey. Beverly Hills, Ca.: Sheridan Supply Services, 1962.

Shanteau, J. Information integration analysis of risky decision making. In M. Kaplan and S. Schwartz (Eds.) Human judgment and decision processes. New York: Academic Press, 1975.

Shanteau, J. and Phelps, R. H. Judgment and swine: Approaches and issues in applied judgment analysis. In M. F. Kaplan and S. Schwartz (Eds.), Human judgment and decision processes: Applications in problem settings, New York: Academic Press 1977.

Shaw, L. and Sichel, H. S. Accident proneness: Research in the occurrence, causation, and prevention of road accidents. New York: Pergamon, 1971.

Shealey, J. E. Risk-taking in skilled task performance. Proceedings of the Eighteenth Annual Meeting of the Human Factors Society. Santa Monica, CA.: Human Factors Society, 1974.

Shelley, M. W., II and Bryan, G. L., (Eds.) Human judgments and optimality, New York: Wiley, 1964.

Slovic, P. Analyzing the expert judge: A descriptive study of a stockbroker's decision processes. Journal of Applied Psychology, 1969 53, 255-263.

Slovic, P. and Lichtenstein, S. Importance of variance preferences in gambling decisions. Journal of Experimental Psychology, 1968, 78, 646-654.

Sommer, R. No, not research. I said evaluation. APA Monitor, 1977, 8, (4), 1-11.

Steiner, I. D. Group process and productivity. New York: Academic Press, 1972.

Schuman, E. A. A conceptual analysis of the accident phenomenon. Social Problems, 1961-60, 8, 241-253.

Strauss, G. and Sayles, L. R. Personnel: The human problems of management. Englewood Cliffs, N. J.: Prentice Hall, 1967.

Tarrants, W. E. Applying measurement concepts to the appraisal of safety performance. Journal of the American Society of Safety Engineers, 1965, 10(5), 15-22.

Taylor, D. H. Accidents, risks, and models of explanation. Human Factors, 1976, 18, 371-380.

Taylor, R. N. and Dunnette, M. D. Influence of dogmatism, risk-taking propensity, and intelligence on decision-making strategies for a sample of industrial managers. Journal of Applied Psychology, 1974, 59, 420-423.

Thorndike, E. L. and Woodworth, R. S. The influence of improvement in one mental function upon the efficiency of other functions. Psychological Review, 1901, 8, 247-261.

Thorpe, J. A., Martin, E. L., Edwards, B. J., and Eddowes, E. E. Situational emergency training: F-15 emergency procedures training program. Williams Air Force Base, Arizona, Technical Report No. AFHRL-TR-76-47(1), June 1976.

Tillman, W. A. and Hobbs, G. E. The accident prone automobile driver: A study of the psychiatric and social background. American Journal of Psychiatry, 1949, 106, 321-331.

Todd, F. J. and Hammon, K. R. Differential feedback in two multiple-cue probabilities learning tasks. Behavioral Science, 1965, 10, 429-435.

Torrance, E. P. and Ziller, R. C. Risk and life experience: Development of a scale for measuring risk-taking tendencies. USAF Personnel Training and Research Center, Research Report 57-23, 1957.

Trans World Airlines. Flight simulator evaluation. Kansas City, Mo.: Trans World Airlines, Flight Operations Training Department, 1969.

Trollip, S. An evaluation of a complex computer-based flight procedures trainer. Savoy, IL.: Aviation Research Laboratory Technical Report No. ARL-77-1/AFOSR-77-1, 1977.

Trollip, S. R. and Ortony, A. Real-time simulation in computer-assisted instruction. Instructional Science, 1977, 6, 135-149.

Trollip, S. and Roscoe, S. N. Computer-assisted instruction in pilot training and certification. Proceedings of the Sixteenth Annual Meeting of the Human Factors Society, Santa Monica, CA.: Human Factors Society, 1972.

Wald, A. Statistical decision functions. New York: Wiley, 1950.

Wallach, M. A. and Kogan, N. Sex differences in judgment processes. Journal of Personality, 1959, 27, 555-564.

Webb, E. J., Campbell, D. T., Schwartz, R. C., and Sechrest, L. Unobtrusive measures. Chicago: Rand McNally, 1966.

Wechsler, D. The measurement of adult intelligence. Baltimore: Williams and Wilkins, 1939.

Williams, A. and Flexman, R. Evaluation of the School Link as an aid in primary flight instruction. University of Illinois Bulletin, 1949, 46, (71), Aeronautics Bulletin 5.

Williges, B. H., Roscoe, S. N. and Williges, R. C. Synthetic flight training revisited. Human Factors, 1973, 15, 549-564.

Wine, J. Test anxiety and direction of attention. Psychology Bulletin, 1971, 76 (2), 92-104.

APPENDIX

Comments from the Aviation Community on Judgment Training and Evaluation

COMMENTS FROM THE AVIATION COMMUNITY ON JUDGMENT TRAINING AND EVALUATION

Introduction

As a part of this contract, information was sought from various sectors of the aviation training community on judgment training and evaluation. Several discussions were held with flight instructors in the Institute of Aviation, University of Illinois at Urbana-Champaign. Other flight training personnel were contacted by phone and personal visits. Organizations contributing information in this way included American Airlines, United Airlines, and the Air Force Flying Training Division at Williams Air Force Base. In addition, written information was sought from nine pilot training organizations representing various levels of training sophistication. Unfortunately, only two of these responded and neither are representative of civil pilot training programs. However, much of what they say is useful to our understanding and application of pilot judgment to Civil Aviation.

The first of these is from Dirk E. Van Dam, Director of the Department of Missionary Aviation of the Moody Bible Institute of Elizabethton, Tennessee. The purpose of the program he describes is to train pilots to fly in remote areas of the world serving missionaries, villages, and local governments. Because of the unique nature of the job for which these pilots are being prepared, a high degree of pilot judgment is required. Accordingly, trainees are selected, trained, and periodically evaluated on the basis of judgment as well as knowledge and skill qualifications.
One part of Moody's selection process is a week-long flight camp in which applicants are given exposure to many aspects of the training program, including some flying. On the basis of performance during this week,

about one-half of the applicants are retained in the program for training over a two-year period.

The second response came from a captain in the United States Air Force who is involved in pilot training. He, and the other seven pilot training persons contacted, were asked to respond to the specific questions given with his comments. Some literature prepared by the Contract Principal Investigator including the proposed judgment definition were also provided as points of reference for these questions.

COMMENTS FROM MOODY AVIATION

Dirk E. Van Dam, Director

Judgment

"The ability to perceive various alternatives in a dynamic situation, and to choose the most acceptable in terms of personal, environmental, and equipment capabilities, while pursuing the desired or planned objective."

Judgment as Criteria for Trainee Selection

It is quite apparent that projection of judgment in flight situations has some parallels in the background life style and record of candidates when we first encounter them. A generally conservative approach to "border-line" situations seems to be an index of whether there is a regard for margin which is so important in flight operations. The way a person thinks about regulations and good operating practices seem to give evidence of this characteristic. In tests that we conduct for personality and behavior patterns, it appears that an ease in decision-making is significantly important, particularly for the pilot candidates. The slow deliberator and the person who is threatened by the "go--no go" type decisions will be often frustrated and sooner yield to other emotional pressures rather than choose the proper alternatives in a situation. I feel quite strongly that a need to be noticed is a characteristic which requires critical attention. The "show-off" tendency may have a decided effect on which option is chosen in flight situations. If this characteristic can be identified and qualified or quantified, it may be supportive of the type of judgment which can be expected. The desired emotional nature of the person probably involves a more restrained or less reactionary response to situations. Discretion for pilots often requires distinguishing between the quick and easy approach, for the plan ahead and preparation approach.

It is probably also true that a certain amount of passivity rather than a defensively aggressive response promotes safety consciousness in an individual. I strongly believe that the person who needs to "prove himself" when faced with questionable situations is a dangerous individual when confronted with hazardous conditions. The person who has a reasonable need to support authority and follow directions, as well as a need for rules and supervision, can be expected to follow standard procedures and good operating practices rather than experiment on the threshold of danger.

Finally, some index of the person's emotional security seems of great importance. The candidate who is "free to fail" in the sense that an error or mistake is treated as an opportunity to learn, has much going for him. In contrast, the person who in any way prefers not to recognize errors or limitations will probably have difficulty in learning, as well as profit little from either his own shortcomings or that of others when brought to his attention. A teachable spirit, in my opinion, is a very important index to the learning which will later take place through personal experience or by vicarious exposure to situations which develop judgment. An individual's personal security about his decision-making ability and the credentials which he recognizes in himself for dealing with flight issues, seems to be an obvious characteristic of those who have long records of safe flight operations.

Initial Indicators of Judgment in Preliminary Training

During the week long process of flight camp evaluation, some of the factors which figure significantly in the overall evaluation, and particularly those which reflect judgment include:

1. Obvious effort and attention to instruction. A casual or careless response is considered negative and usually results in a lack of sensitivity to signals which may affect choices.

2. Relaxation. Alertness without being frustrated or given to anxiety about details without consequence.

3. Division of attention. An ease in dealing with a context where many things need to be integrated and cannot be done "when it's convenient."

4. Response delays. The individual who must "compute" the solution appears to have difficulty in determining alternatives intuitively. Judgment in flight situations seems to require above average ability in thought processes under stress.

5. Confidence. The capacity to realize that the solution <u>is</u> available and that it <u>can</u> be learned.

6. Capacity for problem-solving. The rather obvious awareness and capability that shows even though a procedure has not been prescribed nor instructions specifically given. The situation easily results in a best choice. The "common sense" approach to weighing alternatives and making a choice in line with objectives and definitely within any perimeters required. An obvious positive trait.

7. Initiative. The tendency to take action before a deadline. This quality appears to provide incentive for preparation so that situations which require pre-planning are encountered <u>with</u> that benefit. Initiative seems to be the initial step in planning ahead.

In general, I believe that such qualities as coping with complexity and ambiguity without frustration and a non-rigid approach to the "how" of problem-solving, along with a mild confidence in one's own ability, are some initial indicators of capacity for good judgment in pilot training.

Evidence of Judgment Development

Some of the evidences that practical judgment is being developed through a process of flight training are as follows:

1. Eagerness to learn or high motivation. Punctuality, preparedness, attention to details--perhaps even a little tenseness give evidence that the person will be prepared when a critical decision is to be faced.
2. Teachability. The person is a good listener, asks some questions, is inquisitive and applies what he is told, can accept constructive criticism and suggestions without excuses or defenses. Doesn't always require full explanations to be convinced.
3. Adaptability and Flexibility. The person is able to respond positively to change in procedure or technique. Looks for ways to improve and/or develop his performance. Has interest in ideas and concepts as well as in facts.
4. An intuitive quality in thinking and/or decision-making. Not overly mechanical or academic in treatment of data. Not afraid to experiment within conservative perimeters of personal or equipment limitations. Doesn't require explicit details in order to proceed. Can improvise. Not limited to "computed" answers but can visualize patterns and processes easily. Directs attention to actions which answer the problem; not just explore it.
5. A pattern of good choices. Decisions usually work out to confirm they were the best choice. This person avoids hasty reactions and thinks before speaking or acting. Not impulsive. Generally has alternatives in mind and is not afraid to say "I don't know" or "I'd like to think about that some more." Expects problems but

doesn't worry about things he can't control. Relatively disciplined in some typical areas of life style but not so rigid that he can't do it another way.

6. Applies margin and allowances. Normally has an answer for the question "what if----." Seldom comes up with exact data which is "guaranteed." This person seldom "gets caught" in not having been prepared for assignments, responsibilities and expectations of his peers and supervisors.

Practical Methods of Teaching Judgment

In this area, we lack formal or scientific principles which could be identified with judgmental training. About the best I can provide in the way of general approach is this:

Provide a high exposure to practical situations which require choices. "Live" work and dynamic involvement with operational problems have long been characteristic of Moody training. Apparently, this kind of process removes some of the fear of the unknown and the inhibitions which tend to limit decisions or action. It has often been said that the Moody trained pilot graduates with a lack of confidence. This is probably still true. On the other hand, it seems also apparent that these same people very quickly become confident as they deal with new situations not too unlike those which were part of their training. Rather than practice the well-known, it may be that Moody has developed a process which specializes in the "new and different". This may not generate high confidence initially, but in the long run confidence occurs easily through successful coping. Given the two alternatives of high exposure <u>with</u> the high risk factor involved, or the minimum exposure

with a relatively low risk factor, I believe that operational judgment is best advanced by the former.

In both evaluation and training, the particular manner in which decisions occur may be also somewhat unique in the process developed at Moody. I think we specialize in "first time" decisions and procedures or techniques which can usually be followed without too much revision. There is evidence that procedures and habits learned in initial pilot training are reverted to even many years later when faced with an emergency. Assuming that this is true, it seems vitally important that initial exposure be not only correct in the procedures used; but that it be the kind of exposure which will return as relevant or typical in later years. Simulation has its place for teaching procedure, but the actual experience of having properly recovered may be much more valuable. Critical situations as well as those which are unexpected, when dealt with in an effective manner, produce a high degree of satisfaction and long-term learning in the trainee. I believe that through a process of instructor qualification and standardization, this is characteristic of flight training at Moody.

This training has particular characteristics identified with the above, some of which are:

1. Substantial use of detailed checklist for normal and emergency procedures.
2. Considerable emphasis on critical situations which has been developed through feedback from field operations.
3. A major portion of flight time in tailwheel type aircraft even though this has a higher risk factor, particularly in student training.

4. Over 50% dual instruction in a pilot's total flight time while in the course.
5. Exposure to mountainous terrain as a context of initial training.
6. Exposure to international flying and cross-country experience in the western USA during advanced training.
7. Operational experience in typical flight activity such as ATC operations.

Evaluation of Judgment in Advanced Pilot Training

In observing many trainees over 20+ years, I'm convinced that a significant evidence of judgment is that the person remains a learner throughout his training period. The developing person has a unique balance between confidence through experience, and a discontent because he has not yet fully achieved. It seems apparent that a person who is well aware of his own limitations as well as of his capabilities can make the best judgments in flight situations that are borderline. The learner who continues to have a curiosity about new information on equipment and the environment also continues to be armed with the knowledge that will be necessary. Complacency or overconfidence which become apparent as soon as initial achievements like certificates and ratings occur, is an index of a deterioration in judgment.

The individual who spends much time in "hangar flying" with even mild exaggerations of his performance, can usually be expected to perform poorly in real-life situations. We observe that the individuals who demonstrate the highest quality of flying skill and judgment are also most often the least demonstrative outside the airplane.

Summary

In general, I think it can be rightly observed that pilot judgment is an important element in selection or in encouragement of candidates for flight training. Although a very subjective quality, there appear to be valid tools

by which it can be identified and measured even before pilot training is begun. It appears also realistic that once started, pilot training gives evidence that a person responds favorably or unfavorably to expectations or criteria by which pilot judgment can be evaluated. The training process itself, also can be arranged to provide the necessary emphasis and standards for development of discriminating abilities and response patterns which are expected for the professional pilot level. Those who teach can be prepared to emphasize by theory, demonstration, and directed practice those procedures and guidelines which create a clear and relevant example for trainees. On the negative side, it may be worth consideration that strong indications of behavior or attitudes which identify with poor judgment should result in such candidates being discouraged.

COMMENTS FROM A USAF CAPTAIN
INVOLVED IN PILOT TRAINING

1. What, if anything, would you change in the definition and why?

   I would like to combine the two parts as follows: The ability to search for and establish the relevance of all available information regarding a situation, to specify alternative courses of action, to determine expected outcomes from each alternative, and to then select and execute a suitable course of action within the time frame permitted by the situation. The abilities required to perform the above actions are definable and measurable whereas the motivational actions are not. The motivational aspects I feel have to be assumed. You approach flight instruction from the standpoint that the student wants to learn how to fly and is already motivated to acquire the above abilities. An assumption is made that the student is operating on the esteem or self-actualization level. Our motivational considerations have focused on not doing those things which tend to demotivate the student.

2. Do you agree with the stated goals of pilot judgment? If not, what would you change?

   I agree with the first goal, but not with the second. I don't believe that it is possible to intrinsically motivate a pilot to avoid adverse effects of decision-biasing elements. Attempts to do so in the past relied heavily on the "tried and true" principles of learning, FEAR, RIDICULE AND SARCASM. What is important is that he is cognizant of the laundry list of "things" which impair the decision-making process. For example, don't fly when you are sick, tired, upset, on medication, been drinking, etc., and the physiological/psychological/regulatory why nots. A discussion of these judgement impairing elements should be addressed in a supporting objective of goal number one.

3. On pg. 16, six "decisional activities" are listed.

   Do you feel they would be useful to judgment instruction? Yes.

   What other behaviors which are not listed should be? None.

4. On pg. 17, paragraph 2, we state the need for self discipline on the part of the flight instructor. Is this important or can an instructor say, "Do as I say, not as I do."?

   Yes. To do otherwise would have a demotivating influence.

5. On the bottom of pg. 18, the difficulty of knowing one's own limitations is indicated.

a. How can a flight instructor insure students know them?

By being able to accurately assess his own limitations and express them to his student. In doing so he can establish a relationship between limitations and experience/knowledge. As experience/knowledge levels change, so will limitations.

b. Will this work for these students after they leave the instructional setting?

Possibly! Only accident statistics will tell.

6. What is the best way to insure standardization of minimum judgment competence by all who will fly in our national airspace?

Through standardized flight instruction and reoccurring flight evaluations as evidenced by military and air carrier programs.

7. Would realistic guidelines from the FAA be useful in establishing judgment instruction procedures in your organization?

Yes.